

Characterization of Nonlinear Systems with Memory by Means of Volterra Expansions with Frequency Partitioning: Application to a Cicada Mating Call

Albert H. Nuttall
Adaptive Methods Inc.

Derke R. Hughes
NUWC Division Newport



**Naval Undersea Warfare Center Division
Newport, Rhode Island**

Approved for public release; distribution is unlimited.

20100913314

PREFACE

This report was prepared under Project No. TD 1552, sponsored by the Defense Advanced Research Projects Agency (DARPA, D. Furey).

The technical reviewer for this report was Keith Peters (Code 1513).

Reviewed and Approved: 15 June 2010

A handwritten signature in black ink, appearing to read "David W. Grande". The signature is fluid and cursive, with the first name "David" being the most prominent.

David W. Grande
Head, Sensors and Sonar Systems Department



REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OPM control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY) 15-06-2010		2. REPORT TYPE		3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE Characterization of Nonlinear Systems with Memory by Means of Volterra Expansions with Frequency Partitioning: Application to a Cicada Mating Call				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Albert H. Nuttall Derke R. Hughes				5.d PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Undersea Warfare Center Division 1176 Howell Street Newport, RI 02841-1708				8. PERFORMING ORGANIZATION REPORT NUMBER TR 11,978	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Defense Advanced Research Projects Agency 3701 North Fairfax Drive Arlington, VA 22203				10. SPONSORING/MONITOR'S ACRONYM DARPA	
				11. SPONSORING/MONITORING REPORT NUMBER	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT This report introduces several new concepts to achieve some alleviation of the "curse of dimensionality" in determining the higher order kernels in a Volterra expansion. These concepts include: partitioning of the frequency scale(s), the fundamental region of formation, the diagonal strips in the two-dimensional frequency space, the second- and third-order basis functions, and filtering of the measured nonlinear output to the particular frequency band under investigation. The equations for the real basis functions at second- and third-order take some unexpected forms. These results are based on taking full advantage of symmetry and conjugate-symmetry frequency-domain relations that exist for real, symmetric, second-order and third-order, time-domain kernels. The inability to construct ideal bandpass filters requires the use of a frequency-overlap procedure, followed by discarding of the edge estimates with inherent errors, and retaining only the interior estimates of higher accuracy. Application to a first- and second-order (noise-free) control example with a white broadband excitation gives excellent estimates of all the first- and second-order properties, such as the individual kernels, the individual Volterra waveforms, and the total estimated output waveform. Application to a cicada mating call with a distinctly non-white and non-Gaussian excitation gives good results for the estimated first- and second-order kernels and waveforms, considering the non-optimality of this type of excitation.					
15. SUBJECT TERMS Signal Processing Underwater Acoustics Time-Invariant Nonlinear Systems Volterra Expansions Cicada Calls Frequency Partitioning Wiener Expansions					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT SAR	18. NUMBER OF PAGES 58	19a. NAME OF RESPONSIBLE PERSON Derke R. Hughes
a. REPORT (U)	b. ABSTRACT (U)	c. THIS PAGE (U)			19b. TELEPHONE NUMBER (Include area code) 401-832-5830

TABLE OF CONTENTS

Section	Page
LIST OF ILLUSTRATIONS	ii
1 INTRODUCTION	I
2 BASIC VOLTERRA MODEL	3
3 FIRST-ORDER CONSIDERATIONS	7
3.1 Time-Bandwidth Product Problem	7
3.2 Alleviation of the TF Product at First Order	8
3.3 Sine and Cosine Expansion.....	9
3.4 Key Observation by Means of the Frequency Domain.....	10
3.5 First-Order Waveform Expansion.....	11
3.6 Elimination of Errors at Frequency Joints	12
3.7 Filtering the Nonlinear Response.....	12
4 SECOND-ORDER CONSIDERATIONS	15
4.1 Uniqueness Requirement	15
4.2 Properties of the Second-Order Frequency-Domain Kernel.....	15
4.3 Alternative Form for Second-Order Output $y_2(n\Delta)$	17
4.4 Alternative Derivation of Result (34).....	19
4.5 Basis Functions in Two Dimensions.....	19
4.6 Second-Order Kernel Expansion	20
4.7 Second-Order Waveform Expansion	21
4.8 Discarding of Edge Strips at Second Order	22
4.9 Test Procedure for a First-Order and Second-Order Example.....	22
4.10 Interpretation and Usefulness	24
5 SECOND-ORDER NUMERICAL RESULTS FOR THE CONTROL EXAMPLE	25
6 THIRD-ORDER CONSIDERATIONS	29
6.1 Uniqueness Requirement	29
6.2 Properties of the Third-Order Frequency-Domain Kernel.....	29
6.3 Alternative Form For Third-Order Output $y_3(n\Delta)$	30
6.4 Basis Functions in Three Dimensions.....	31
6.5 Third-Order Kernel Expansion	33
6.6 Third-Order Waveform Expansion	33
7 APPLICATION TO A CICADA SONG	35
7.1 Waveforms and Spectra	35
7.2 First-Order Fit	40
7.3 Second-Order Idealizations.....	40

TABLE OF CONTENTS (Cont'd)

Section	Page
8 SECOND-ORDER, TWO-INPUT, ONE-OUTPUT VOLTERRA MODEL.....	51
9 SUMMARY	53

LIST OF ILLUSTRATIONS

Figure	Page
1 Measurement and Fitting Procedure	4
2 Diagonal Strips in f_1, f_2 Plane.....	18
3 Test Procedure for Second Order.....	23
4 First-Order Kernels $h_{11}(b)$ $h_1(k)$ $e_{h1}(r)$	26
5 Exact Second-Order Kernel $h_{22}(k_1, k_2)$	26
6 Model Second-Order Kernel $h_2(k_1, k_2)$	27
7 First-Order Waveforms $z_1(b)$ $y_1(k)$ $e_1(r)$	27
8 Second-Order Waveforms $z_2(b)$ $y_2(k)$ $e_2(r)$	28
9 Waveforms $z(b)$ $y(k)$ $e(r)$	28
10 Waveform L_1	36
11 Waveform L_2	36
12 Waveform M	37
13 Waveform L_2 (0.05-Second Segment)	37
14 Waveform L_2 (0.01-Second Segment)	38
15 Spectrum of Waveform L_1	38
16 Spectrum of Waveform L_2	39
17 Spectrum of Waveform M	39
18 First-Order Kernel $h_1(k)$, $ h_1(k) $ (:), and Envelope (r).....	41
19 First-Order Kernel H_1	41
20 Smoothed Second-Order Kernel $ h_2 $ (Observation Elevation Angle 35°)	43
21 Smoothed Second-Order Kernel $ h_2 $ (Observation Elevation Angle 90°)	43
22 Second-Order Kernel H_2 (Observation Elevation Angle 50°)	45
23 Second-Order Kernel H_2 (Observation Elevation Angle 90°)	45
24 First-Order Waveform y_1	46
25 First-Order Spectrum Y_1	46
26 Second-Order Waveform y_2	48
27 Second-Order Spectrum Y_2	48
28 Total Waveforms $z(k)$ $y(r)$	49
29 Waveforms $z(k)$ $y_1(b)$ $y_2(r)$ (Laser L_2 of Cicada Song S14).....	49
30 Waveforms $z(k)$ $y_1(b)$ $y_2(r)$ (Laser L_1 of Cicada Song S9).....	50
31 Waveforms $z(k)$ $y_1(b)$ $y_2(r)$ (Laser L_2 of Cicada Song S9).....	50
32 Second-Order, Two-Input, One-Output Volterra Model	52

CHARACTERIZATION OF NONLINEAR SYSTEMS WITH MEMORY BY MEANS OF VOLTERRA EXPANSIONS WITH FREQUENCY PARTITIONING: APPLICATION TO A CICADA MATING CALL

1. INTRODUCTION

The Wiener characterization of time-invariant nonlinear systems with memory utilizes an expansion involving kernels of various orders, typically limited to third order or less in practice. This limitation to third order is due to the “curse of dimensionality,” that is, the fact that the required number of kernel coefficients increases exponentially with the order of fit adopted in the expansion. Also, great care must be taken in keeping the condition number of the very large relevant data matrix within reasonable bounds, so that its pseudo-inverse and Wiener expansion are accurate. This must be achieved even when the Gaussian excitation has a colored spectrum and, therefore, requires a special type of expansion of the kernel terms themselves. The end result of the characterization is a *quantitative* statement and breakdown of the exact amounts of each type of nonlinearity at the nonlinear system output. This information has obvious applications, including the underwater acoustic channel, by indicating the strength of the system nonlinearities and whether they can be ignored or whether they should be exploited for improved detection and/or classification purposes.

However, there are numerous instances where the excitation $x(t)$ of the nonlinear system under investigation is not under one’s control, but is instead dictated by the particular physical situation at hand. Such a situation arises in the case of a physical cicada mating call, where the excitation is the mechanical movement of the creature’s tymbal(s), and the nonlinear system response $z(t)$ is the audio pressure waveform received at a recording device some distance away. In such a case, appeal *cannot* be made to the Wiener expansion because the higher order moments of excitation $x(t)$ are then unknown. Whereas all the odd-order moments of a stationary, zero-mean, Gaussian process are zero, and all the even-order moments are expressible in terms of just the correlation function $R_x(\tau)$ of the Gaussian process $x(t)$, *none* of the higher order moments of a non-Gaussian process are generally known, and are extremely difficult to estimate due to the need for an extreme amount of data about the excitation process $x(t)$. For this reason, in the current cicada application, it is necessary to abandon the Wiener procedure for third order and above and, instead, revert to the Volterra expansion, which makes no assumptions about the statistics of the excitation $x(t)$.

In an attempt to alleviate the curse of dimensionality, a frequency partitioning scheme was proposed in an earlier technical report.* However, it has been found that the non-overlapping square regions there, in the two-dimensional frequency space f_1, f_2 , interact with each other, cross-contaminating each estimate. Also, off-diagonal frequency squares bring up the question

*Albert H. Nuttall, “Characterization of Nonlinear Systems with Memory, by Means of Volterra Expansions,” Adaptive Methods Inc., Middletown, RI, 30 December 2009.

as to what frequency band to fit, the lower or upper or both. These problems have led to a very different partitioning of the two-dimensional frequency space, to be described here, that solves both problems and yields kernel and waveform estimates of both first and second order that are excellent for the control example used to test this new technique. Numerical results are presented to demonstrate the efficiency of this latest approach. An application to a cicada mating call is made that illustrates a very interesting second-order behavior.

2. BASIC VOLTERRA MODEL

Nonlinear system excitation $x(t)$ is sampled at frequency f_s Hz, resulting in time-sampling increment $\Delta = 1/f_s$ seconds and sampled sequence $\{x(n\Delta)\}$. For simplicity of notation, the Δ symbol will frequently be suppressed and the excitation sequence will be denoted simply by $\{x(n)\}$. At other times, Δ will be kept in order to stress the time dependence. Also, sequence $\{x(n)\}$ will frequently be referred to as a waveform.

Consider a time-invariant nonlinear system with actual sampled input sequence $\{x(n)\}$ and actual sampled output sequence $\{z(n)\}$, both of which are sampled at the same rate f_s and recorded simultaneously. The causal time-invariant Volterra *model* sampled output sequence $\{y(n)\}$ is then given, to third order, by

$$\begin{aligned} y(n) = & h_0 + \sum_{k_1=0}^{K-1} h_1(k_1) x(n-k_1) + \sum_{k_1=0}^{K-1} \sum_{k_2=0}^{K-1} h_2(k_1, k_2) x(n-k_1) x(n-k_2) \\ & + \sum_{k_1=0}^{K-1} \sum_{k_2=0}^{K-1} \sum_{k_3=0}^{K-1} h_3(k_1, k_2, k_3) x(n-k_1) x(n-k_2) x(n-k_3) \\ \equiv & y_0 + y_1(n) + y_2(n) + y_3(n), \end{aligned} \quad (1)$$

where h_0, h_1, h_2, h_3 are the zeroth-order through third-order (time-invariant) time-domain kernels of the Volterra expansion. It is assumed that the Volterra kernels h_1, h_2, h_3 are represented with the same time-sampling increment Δ as used for the nonlinear system input and output waveforms $x(n)$ and $z(n)$. (The $\{ \}$ symbol around a sequence will be dropped from here on.) It is also assumed that the same “memory length” K in equation (1) is appropriate for all three orders of these kernels. This may not be the case in practice; if not, different sizes K_1, K_2, K_3 of the summations in equation (1) may be required for a decent approximation of nonlinear system output $z(n)$ by the model output $y(n)$.

Although the input data $x(n)$ appear in a *nonlinear* fashion in component model outputs $y_2(n)$ and $y_3(n)$ in equation (1), sequence $x(n)$ is *known*, and the required second-order and third-order input-data products in equation (1) can be easily calculated and stored. The *unknowns* in the Volterra expansion (1) are the four kernels h_0, h_1, h_2, h_3 , which appear *linearly* in the model output $y(n)$. This observation strongly suggests the use of a least squares approach in attempting to fit model output $y(n)$ to the actual measured nonlinear system output $z(n)$, through optimum choice of these Volterra kernels; see figure 1. Namely, the equations determining the best kernels h_0, h_1, h_2, h_3 will be the solutions of simultaneous *linear* equations under the least squares philosophy. This is a very important consideration.

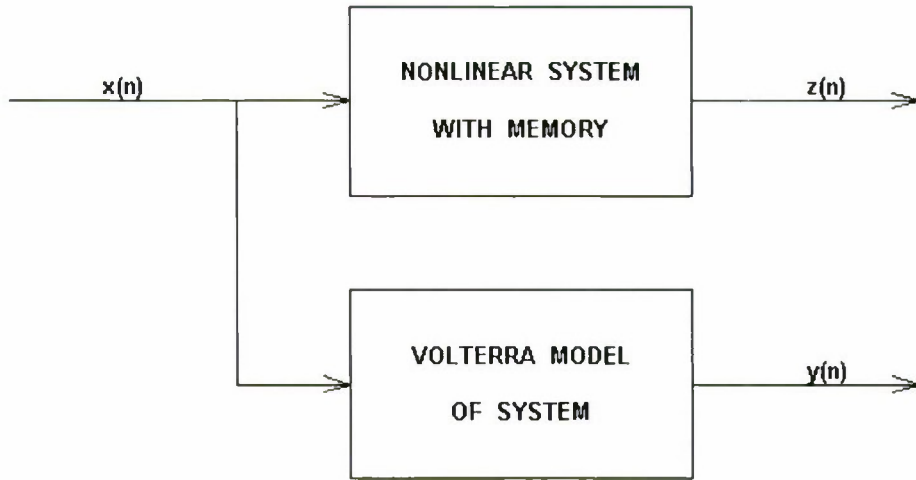


Figure 1. Measurement and Fitting Procedure

One major problem with this Volterra approach is the ill-conditioning of the very large data matrix that must be pseudo-inverted in the least squares processing approach. For a Gaussian excitation $x(n)$, this ill-conditioning problem could be partially overcome by switching to a Wiener expansion with uncorrelated components (on an ensemble average basis); but that route is not available for third order and above in the current case of a non-Gaussian excitation $x(n)$. However, a second-order Wiener modification *is still applicable* and is described below.

The additional problem of very large data matrices can be overcome by time-partitioning of the excitation $x(n)$, accompanied by temporary storage on a hard drive. When particular sub-matrices are needed for computation in random access memory (RAM), these matrices can be recalled as needed from the hard drive. Of course, the tradeoff here is that the execution time can increase considerably.

However, the really major problem associated with both the Volterra and the Wiener expansions is the curse of dimensionality (COD), namely, the extreme number of coefficients (kernel values) required in equation (1). At first order, the number of coefficients that must be determined is $M_1 = K$; at second order, the number of coefficients is approximately $M_2 = K^2/2$; and at third order, it is approximately $M_3 = K^3/6$. The helpful denominator factors of 2 and 6 come about through the judicious imposition of symmetry on the unknown kernels h_2 and h_3 , *without* loss of generality.

The programming required for these tasks requires careful consideration in order to avoid ill-conditioning, storage limitations, and excessive execution time. Once available, the resulting program(s) can be utilized to determine the amounts and orders of nonlinearity in the cicada

ming call. If there is significant energy in the higher order nonlinear terms, this will indicate quantitatively just how important it is to determine exactly how the cicada can emit such strong signals from its small body. Practical applications of this information are of obvious importance in many fields, including underwater acoustics.

For an expansion limited to second order, the Wiener modification takes the form

$$\begin{aligned}
 y(n) &= w_0 + \sum_{k_1=0}^{K-1} w_1(k_1) x(n-k_1) + \sum_{k_1=0}^{K-1} \sum_{k_2=0}^{K-1} w_2(k_1, k_2) [x(n-k_1) x(n-k_2) - R_x(k_1 - k_2)] \\
 &= \left[w_0 - \sum_{k_1=0}^{K-1} \sum_{k_2=0}^{K-1} w_2(k_1, k_2) R_x(k_1 - k_2) \right] + \sum_{k_1=0}^{K-1} w_1(k_1) x(n-k_1) \\
 &\quad + \sum_{k_1=0}^{K-1} \sum_{k_2=0}^{K-1} w_2(k_1, k_2) x(n-k_1) x(n-k_2) \\
 &\equiv y_0 + y_1(n) + y_2(n),
 \end{aligned} \tag{2}$$

where w_0, w_1, w_2 are the Wiener kernels. The subtraction of the excitation correlation function $R_x(k_1 - k_2)$ in the first line of equation (2) makes the second-order (double-summation) term in the top line uncorrelated with the zeroth-order term w_0 , *regardless* of the statistics of excitation $x(n)$. If $x(n)$ has zero mean, then the zeroth-order term and the first-order term in the first line of equation (2) are also uncorrelated with each other. But that still leaves the cross-correlation between the first-order term and the second-order term unknown in the case of a non-Gaussian excitation $x(n)$. Nevertheless, this Wiener modification in equation (2) is well worth considering and trying, when the expansion is limited to second order. It can lead to significantly smaller condition numbers for the large data matrix involved in the least squares procedure. Only the correlation function $R_x(k)$ of the excitation $x(n)$ needs to be estimated from the available data; this is a very simple and quickly done task.

The interrelationships between the Volterra and Wiener kernels at second order are seen, from equations (1) and (2), to be

$$h_0 = w_0 - \sum_{k_1=0}^{K-1} \sum_{k_2=0}^{K-1} w_2(k_1, k_2) R_x(k_1 - k_2), \quad h_1 = w_1, \quad h_2 = w_2. \tag{3}$$

Thus, if the first line of Wiener model (2) is used to least squares fit the measured nonlinear output $z(n)$, the corresponding Volterra kernels can then be determined from equation (3), and the corresponding Volterra waveforms can be determined from the last two lines of equation (2).

For completeness, the Wiener modification at third order for a Gaussian excitation would be

$$\begin{aligned}
 y(n) = & w_0 + \sum_{k_1=0}^{K-1} w_1(k_1) x(n-k_1) + \sum_{k_1=0}^{K-1} \sum_{k_2=0}^{K-1} w_2(k_1, k_2) [x(n-k_1) x(n-k_2) - R_x(k_1 - k_2)] \\
 & + \sum_{k_1=0}^{K-1} \sum_{k_2=0}^{K-1} \sum_{k_3=0}^{K-1} w_3(k_1, k_2, k_3) [x(n-k_1) x(n-k_2) x(n-k_3) - 3 R_x(k_2 - k_3) x(n-k_1)],
 \end{aligned} \tag{4}$$

using the assumed symmetry of Wiener kernel $w_3(k_1, k_2, k_3)$, *without* loss of generality. The corresponding Volterra kernels are then given by

$$\begin{aligned}
 h_0 &= w_0 - \sum_{k_1=0}^{K-1} \sum_{k_2=0}^{K-1} w_2(k_1, k_2) R_x(k_1 - k_2), \\
 h_1(k_1) &= w_1(k_1) - 3 \sum_{k_2=0}^{K-1} \sum_{k_3=0}^{K-1} w_3(k_1, k_2, k_3) R_x(k_2 - k_3), \\
 h_2 &= w_2, \quad h_3 = w_3,
 \end{aligned} \tag{5}$$

and would be computed *after* the least squares fit of Wiener model (4) to measured nonlinear system output data $z(n)$ was completed. Volterra waveforms $y_0, y_1(n), y_2(n), y_3(n)$ are still given by equation (1) in terms of the Volterra kernels.

3. FIRST-ORDER CONSIDERATIONS

3.1 TIME-BANDWIDTH PRODUCT PROBLEM

Corresponding to the first-order time-domain description $h_1(k) \equiv h_1(k_1\Delta)$ in equation (1), it is useful to consider the equivalent frequency-domain descriptor, namely, the first-order frequency-domain kernel (discrete-time Fourier transform)

$$H_1(f_1) = \sum_{k_1=0}^{K-1} h_1(k_1\Delta) \exp(-i 2 \pi k_1 \Delta f_1). \quad (6)$$

This relation can be inverted by means of the inverse Fourier transform to obtain

$$h_1(k_1\Delta) = \Delta \int_{-F}^F df_1 H_1(f_1) \exp(i 2 \pi k_1 \Delta f_1), \text{ for } k_1 = 0 : K - 1, \quad (7)$$

where the assumed band-limited character of the frequency-domain kernel $H_1(f_1)$ has been explicitly indicated as $(-F, F)$. That is, $F < 1/(2\Delta) = f_s/2$, the Nyquist frequency.

Suppose that first-order time-domain kernel $h_1(k_1\Delta)$ in equations (1) and (6) has a time extent of T seconds,

$$T = K \Delta, \quad (8)$$

and that the first-order frequency-domain kernel $H_1(f_1)$ in equation (6) has a bandwidth (frequency extent) of F Hz, assumed lowpass about zero frequency, as indicated in equation (7). Then, to get a *complete* characterization of the continuous function $h_1(\tau)$, it is necessary to sample $h_1(\tau)$ with time increment $1/(2F)$ seconds or finer. Since the time duration of $h_1(\tau)$ is T seconds, a total of at least $2TF$ samples must be employed in order not to miss any significant information about $h_1(\tau)$. This implies that the size K in equation (1) must be at least $2TF$.

Although this value $K = 2TF$ is usually manageable at first order, as far as least squares is concerned, the numbers of required coefficients for specification at second and third orders, given by

$$M_2 = (2TF)^2/2 \quad \text{and} \quad M_3 = (2TF)^3/6, \quad (9)$$

respectively, can quickly get out of hand. For example, if $2TF = 100$, then

$$M_1 = K = 100, \quad M_2 = 5,000, \quad \text{and} \quad M_3 = 167,000. \quad (10)$$

In the normal equations that arise in least squares, the size of the data product matrix that must be inverted is $M \times M$. The $M_2 \times M_2$ case can often be solved with current-day computer RAM, but the $M_3 \times M_3$ matrix will often not fit into RAM. If a simultaneous fit of *all* the components in equation (1) to measured nonlinear system output $z(n)$ were of interest, it can happen that nothing at all could be achieved, because of excessive storage requirements. Some method of addressing *partitions* of the various kernels is needed if meaningful useful estimates of the various kernels are to be obtained at higher orders.

3.2 ALLEVIATION OF THE *TF* PRODUCT AT FIRST ORDER

To see how a large *TF* product can be alleviated, recall equation (7). For a real first-order kernel $h_1(k_1\Delta)$, develop it as

$$\begin{aligned} h_1(k_1\Delta) &= 2\Delta \operatorname{Re} \int_0^F df_1 H_1(f_1) \exp(i2\pi k_1\Delta f_1) \\ &= 2\Delta \operatorname{Re} \int_0^W df_1 H_1(f_1) \exp(i2\pi k_1\Delta f_1) + 2\Delta \operatorname{Re} \int_W^{2W} df_1 H_1(f_1) \exp(i2\pi k_1\Delta f_1) + \cdots \quad (11) \\ &\equiv h_{10}(k_1\Delta) + h_{11}(k_1\Delta) + h_{12}(k_1\Delta) + \cdots + h_{1J}(k_1\Delta), \quad J = F/W - 1. \end{aligned}$$

Subscript $1j$ denotes the j -th band of the first-order kernel. This breakdown of kernel $h_1(k_1\Delta)$ is not an approximation; it is exact. The frequency content of component kernel $h_{10}(k_1\Delta)$, namely, its frequency-domain kernel $H_{10}(f_1)$, is limited to positive frequency band $(0, W)$, while $H_{1j}(f_1)$ for component kernel $h_{1j}(k_1\Delta)$ is limited to positive frequency band $(jW, (j+1)W)$ for $j = 0 : J$. This reduces the time-bandwidth product for each component kernel by a factor of F/W . On the other hand, there are now $J + 1 = F/W$ narrowband kernels to be determined instead of just one broadband kernel. If one can evaluate (estimate) all the individual kernels $h_{1j}(k_1\Delta)$, $j = 0 : J$, they can be added together to yield the total first-order time-domain kernel $h_1(k_1\Delta)$, and thereby determine the *entire* first-order model output $y_1(n)$ in equation (1).

The first-order model output sequence $y_1(n)$ in equation (1) can now be expressed as the sum of $J + 1$ narrowband components:

$$y_1(n) = y_{10}(n) + y_{11}(n) + \cdots + y_{1J}(n), \quad (12)$$

where

$$y_{1j}(n) = \sum_{k_1=0}^{K-1} h_{1j}(k_1) x(n - k_1) = \Delta \int df_1 H_{1j}(f_1) X(f_1) \exp(i2\pi f_1 n \Delta), \quad (13)$$

and where

$$X(f) = \sum_n \exp(-i 2 \pi f n \Delta) x(n\Delta) \quad (14)$$

is the Fourier transform of the discrete-time excitation $x(n\Delta)$. Regardless of the spectral content $X(f)$ of time-excitation $x(n\Delta)$, a band-limited, first-order, frequency-domain kernel $H_{1j}(f)$ will limit the spectral content of model output component $y_{1j}(n)$ to the frequency band $(jW, (j+1)W)$. If this particular model component $y_{1j}(n)$ is to be *separately* fitted to the measured system output $z(n)$, it makes sense to *first filter* the nonlinear system output $z(n)$ to this same frequency band under consideration. Call this filtered version $z_j(n)$ for the j -th band, $j = 0 : J$.

If this pre-filtering of $z(n)$ is not conducted, the least squares fit of bandpass sequence $y_{1j}(n)$ to the original nonlinear system output sequence $z(n)$ will result in the least squares procedure attempting to match out-of-band frequency components, which will be considered as noise or outliers to be fitted as well as possible. Therefore, it is mandatory to remove these contaminating out-of-band components in $z(n)$ *before* attempting a least squares fit of $y_{1j}(n)$ to the available system output data. This eliminates out-of-band frequency components in $z(n)$ that the model components can't possibly fit because of their intentionally limited frequency content. Even if this partitioning of the frequency scale were not adopted at first order, nonlinearity output process $z(n)$ should still be filtered to the model's band of interest before doing any least squares fitting, because a model can't match frequency components outside its own band.

The breakdown of the first-order time-domain kernel in equation (11) is usually not necessary in practice because the TF product is typically not that large, and the corresponding least squares data matrix can be easily fitted into RAM. The real benefit of reducing the TF product is realized much more strongly when attempting second- and third-order least squares fits.

3.3 SINE AND COSINE EXPANSION

For the first-order kernel (waveform) $h_{10}(\tau)$ with duration T and positive-frequency extent $(0, W)$, an attractive expansion (including negative frequencies) is afforded by the standard trigonometric form

$$h_{10}(k_1\Delta) = \sum_{m=0}^M a(m) \cos\left(2\pi \frac{m}{T} k_1\Delta\right) + \sum_{m=1}^M b(m) \sin\left(2\pi \frac{m}{T} k_1\Delta\right), \quad (15)$$

where $M/T = W$, the highest frequency in this particular low-frequency band, and the $\{a(m)\}$ and $\{b(m)\}$ coefficient sequences are real. The number of coefficients to be determined is about

$2M = 2TW$, not $2TF$. Thus, the frequency-partitioning procedure achieves a worthwhile reduction in the sizes of the individual fitting procedures. There is no $b(0)$ term in this usual harmonic expansion in equation (15); this is a trivial point at first-order, but will become significant at second-order. Similar harmonic expansions to equation (15) hold for the frequency bands $(W, 2W)$, $(2W, 3W)$, ..., $(F - W, F)$.

The first-order time-domain kernel $h_1(k_1\Delta)$ is forced to be real. Therefore, the first-order frequency-domain kernel $H_1(f_1)$ defined in equation (6) satisfies a conjugate symmetry property:

$$H_1(-f_1) = H_1(f_1)^*. \quad (16)$$

If one specifies $H_1(f_1)$ for positive frequencies f_1 , its values for negative frequencies are automatically fixed. This positive frequency range is called the “fundamental region” at first order.

A single positive-frequency elemental component of $h_1(k_1\Delta)$ in this fundamental region is

$$[a(m) - i b(m)] \exp\left(i 2\pi \frac{m}{T} k_1 \Delta\right), \quad a(m), b(m) \text{ real}; \quad f_m = \frac{m}{T} \geq 0. \quad (17)$$

Upon addition of the conjugate contribution from the corresponding negative frequency $-m/T$, the fundamental *real* basis function for this single frequency f_m is, after scaling by $1/2$,

$$a(m) \cos\left(2\pi \frac{m}{T} k_1 \Delta\right) + b(m) \sin\left(2\pi \frac{m}{T} k_1 \Delta\right). \quad (18)$$

This is, of course, the same expression as already given in equation (15); however, this latter development serves to introduce the “fundamental region” concept, which will take on a much greater significance at second and higher orders. The basis functions are simply cosines and sines at first order.

3.4 KEY OBSERVATION BY MEANS OF THE FREQUENCY DOMAIN

The first-order component output time waveform $y_{1j}(n\Delta)$ in equations (12) and (13) was expressed in terms of frequency-domain quantities as

$$y_{1j}(n\Delta) = \Delta \int df \exp(i 2\pi f n\Delta) H_{1j}(f) X(f), \quad (19)$$

where the excitation spectrum $X(f)$ is generally broadband. The *only* place where time variable $n\Delta$ appears on the right-hand side of this expression is inside the $\exp(\)$ with the multiplier f . If waveform $y_{1j}(n\Delta)$ on the left-hand side of equation (19) is to contain frequency

components only in frequency band $(jW, (j+1)W)$, then frequency-domain kernel $H_{1j}(f)$ must be nonzero *only* in this same band. That is, one must have $H_{1j}(f)$ nonzero only for

$$jW < f < (j+1)W \quad (20)$$

and the corresponding negative frequencies. This appears to be a trivial point here at first order; however, this observation will be of crucial importance at second and higher orders.

3.5 FIRST-ORDER WAVEFORM EXPANSION

When the trigonometric expansion for $h_{10}(k_1\Delta)$ in equation (15) is substituted into the first equality on the left-hand side of equation (13), and the summations are interchanged, the result is

$$y_{10}(n\Delta) = \sum_{m=0}^M a(m) x_c(n\Delta, m) + \sum_{m=1}^M b(m) x_s(n\Delta, m), \quad (21)$$

where

$$\begin{aligned} x_c(n\Delta, m) &= \sum_{k=0}^{K-1} \cos\left(2\pi \frac{m}{T} k\Delta\right) x(n\Delta - k\Delta), \\ x_s(n\Delta, m) &= \sum_{k=0}^{K-1} \sin\left(2\pi \frac{m}{T} k\Delta\right) x(n\Delta - k\Delta). \end{aligned} \quad (22)$$

These latter summations are convolutions of the trigonometric basis functions with the excitation $x(n\Delta)$. These waveform sequences $\{x_c(n\Delta, m)\}$ and $\{x_s(n\Delta, m)\}$ can be precomputed for any m values of interest and stored. According to the summation limit in equation (21), these waveforms are approximately band limited to $(0, W)$ Hz, as desired, since $M/T = W$. Expansions similar to equation (21) can be developed for the other frequency bands $(W, 2W), \dots, (F - W, F)$.

The expansion of first-order component waveform $y_{10}(n\Delta)$ in equation (21) is linear in the unknown real coefficients $a(m)$ and $b(m)$. These latter coefficients must now be chosen so that $y_{10}(n\Delta)$ approximates $z_0(n\Delta)$ as closely as possible in a least squares sense; see the discussion following equation (13). The result of the least squares minimization will be a set of $2M - 1$ simultaneous linear equations for the unknown real coefficients. (There will be $2M$ simultaneous linear equations for the real coefficients in all the other bands, $j = 1 : J$.)

Once the coefficients $a(m)$ and $b(m)$ have been determined from the least squares fit to $z_0(n\Delta)$, equation (15) is used to find the estimated kernel $h_{10}(k_1\Delta)$. Also, equations (21) and (22) then yield the best-fitting waveform $y_{10}(n\Delta)$ for the $(0, W)$ frequency band. This fitting procedure must be repeated for all the other W -wide bands up to frequency F .

After accomplishing the fits for all the frequency bands up to frequency F , producing fitted first-order waveforms $y_{10}(n\Delta), y_{11}(n\Delta), \dots, y_{1J}(n\Delta)$, these waveforms are added together to get the total first-order fit $y_1(n\Delta)$ to the *total* nonlinear broadband measured response $z(n\Delta)$. Then, a total error waveform at first order can be computed according to

$$e_1(n\Delta) = y_1(n\Delta) - z(n\Delta). \quad (23)$$

3.6 ELIMINATION OF ERRORS AT FREQUENCY JOINTS

It has been found that the spectrum of error $e_1(n\Delta)$ in equation (23) has peaks in the neighborhoods of frequencies $W, 2W, 3W, \dots$. These latter frequencies are the locations of the “joints” of the frequency partitioning in equation (11). The reason for the errors at these frequency joints is the inability to do perfect bandpass filtering on output process $z(n\Delta)$. Every physical filter has to have a transition region in frequency, from its full response to its maximum attenuation level. Also, the finite-duration (gated) sines and cosines employed in expansion (15) always have side lobes in the frequency domain and will therefore have some nonzero response to frequency components outside their nominal “ideal” band.

A method for overcoming this problem is best explained by taking a numerical example. Let bandwidth $W = 6$ kHz. First, perform a least squares fit over (0:6) kHz. Retain the coefficients $a(m)$ and $b(m)$ *only* corresponding to the band (0:5) kHz. Discard the edge coefficients corresponding to (5:6) kHz, because they are expected to be in error.

Next, perform a separate fit over the (4:10)-kHz band. Retain only the coefficients pertaining to the interior (5:9)-kHz band. Discard both edge coefficients corresponding to bands (4:5) kHz and (9:10) kHz. Next, fit over (8:14) kHz. Retain only the coefficients for (9:13) kHz. The general procedure is now obvious.

In this manner, although 6 kHz is fit at a time, only the interior 4-kHz coefficients are retained, because these are the coefficients expected to be correct. The overlaps in frequency, for each successive fit, are necessary in order to be able to discard edge coefficients that are not considered trustworthy.

3.7 FILTERING THE NONLINEAR RESPONSE

By design, component waveform $y_{10}(n\Delta)$ has frequency content only in the frequency band $(0, W)$; of course, this is only approximately true, as noted above. But nonlinear system output $z(n\Delta)$ is very broadband. As discussed in the sequel to equation (13), model output $y_{10}(n\Delta)$ cannot possibly match the frequency components of $z(n\Delta)$ that are outside the band $(0, W)$, whether they are signal or noise components. On the other hand, least squares is a greedy

procedure and will inherently *attempt* to fit out-of-band components to some extent. Therefore, filtering of $z(n\Delta)$ is necessary before attempting any least squares fits.

In doing the bandpass filtering of nonlinearity response $z(n\Delta)$ to the W -wide band of the current (j -th) fit, no distortion of the in-band components of $z_j(n\Delta)$ can be allowed to occur. Therefore, it is necessary to utilize a bandpass filter with a *linear* phase shift over the central 4 kHz of each sub-band of width 6 kHz. A symmetric finite-impulse-response (FIR) filter fits this requirement perfectly. In fact, the phase response of such a filter is linear for *all* frequencies. Also, the amplitude response of each filter must be extremely flat over the interior 4 kHz of its passband. A couple of computer routines that meet these requirements admirably are MATLAB's `firls` and `firpm` routines, where the `ls` stands for least squares and the `pm` stands for Parks-McClellan. A 4-kHz passband with 1-kHz transition bands at each edge is easily achieved by either of these routines.

For a linear filter with causal impulse response duration L seconds, the transient response is of length L seconds; this startup transient at the filter output must be discarded, as it cannot be fitted properly. In addition, the time delay of the filtered output $z_j(n\Delta)$ is $L/2$ seconds, not L seconds; accordingly, excitation $x(n\Delta)$ must be delayed by this same $L/2$ amount before attempting any fits.

4. SECOND-ORDER CONSIDERATIONS

4.1 UNIQUENESS REQUIREMENT

The second-order Volterra component waveform was given in equation (1) as

$$y_2(n) = \sum_{k_1=0}^{K-1} \sum_{k_2=0}^{K-1} h_2(k_1, k_2) x(n-k_1) x(n-k_2). \quad (24)$$

It must be noted immediately that there can be no unique solution for the second-order time-domain kernel $h_2(k_1, k_2)$. Only the *sum*

$$h_2(k_1, k_2) + h_2(k_2, k_1) \quad (25)$$

of symmetrically located kernel values (about the 45° line in k_1, k_2 space) affects output $y_2(n)$. This is due to the symmetry of the product of the two x values in variables k_1, k_2 in equation (24). To enable a unique solution for the second-order time-domain kernel, it is necessary to impose some condition. The condition adopted here is to take the second-order kernel to be real and symmetric:

$$h_2(k_2, k_1) = h_2(k_1, k_2). \quad (26)$$

There is *no* loss of generality in imposing this symmetry relation on the second-order time-domain kernel $h_2(k_1, k_2)$, because the output $y_2(n)$ of equation (24) depends only on the sum in equation (25), no matter what the excitation $x(n)$ is. Also, significant advantage will be taken of this symmetry property, including a reduction in the number of unknown coefficients that must be determined, as well as avoidance of singular matrices.

4.2 PROPERTIES OF THE SECOND-ORDER FREQUENCY-DOMAIN KERNEL

The second-order frequency-domain (complex) kernel corresponding to $h_2(k_1\Delta, k_2\Delta)$ is

$$H_2(f_1, f_2) = \sum_{k_1=0}^{K-1} \sum_{k_2=0}^{K-1} h_2(k_1\Delta, k_2\Delta) \exp(-i2\pi f_1 k_1\Delta - i2\pi f_2 k_2\Delta). \quad (27)$$

The inverse relation is

$$h_2(k_1\Delta, k_2\Delta) = \Delta^2 \iint df_1 df_2 H_2(f_1, f_2) \exp(i2\pi f_1 k_1\Delta + i2\pi f_2 k_2\Delta). \quad (28)$$

Let the time duration of $h_2(k_1\Delta, k_2\Delta)$ in each time dimension be T seconds, and let the frequency extent of $H_2(f_1, f_2)$ in each frequency dimension be $(-F, F)$ Hz. These limits are identical to those used for the first-order situation. Then, using symmetry, approximately $(2TF)^2/2$ samples are required to completely characterize the real second-order time-domain kernel $h_2(k_1\Delta, k_2\Delta)$. This rapid increase in the number of required coefficients for characterization, in progressing from first order to second order, is the COD.

Since $h_2(k_1\Delta, k_2\Delta)$ is forced to be real, it follows immediately from equation (27) that

$$H_2(-f_1, -f_2) = H_2(f_1, f_2)^*. \quad (29)$$

And, from equations (26) and (27), there follows

$$H_2(f_2, f_1) = H_2(f_1, f_2). \quad (30)$$

That is, the complex second-order frequency-domain kernel $H_2(f_1, f_2)$ satisfies both a symmetry relation about the $+45^\circ$ line in f_1, f_2 space, and a conjugate symmetry relation through the origin of f_1, f_2 space.

These symmetry properties mean that complex frequency-domain kernel $H_2(f_1, f_2)$ needs to be specified only in a 90° sector of f_1, f_2 space; it is then automatically fixed in the rest of the two-dimensional frequency plane. In particular, the “east” sector, composed of the 90° sector centered on the positive f_1 axis, is adopted as the “fundamental region” in f_1, f_2 space. Mathematically, that is the region

$$f_1 \geq 0, \quad -f_1 \leq f_2 \leq f_1. \quad (31)$$

Advantage must always be taken of these symmetry properties because they afford a significant reduction in the number of unknowns that must be determined (estimated). Only the values of $H_2(f_1, f_2)$ in the fundamental region (31) need to be specified.

The area of the complete frequency space f_1, f_2 , of extent $(-F, F)$ in each dimension, is $(2F)^2$. On the other hand, the area covered by equation (31) is only

$$\int_0^F df_1 \int_{-f_1}^{f_1} df_2 = F^2 = \frac{1}{2} \frac{1}{2} (2F)^2. \quad (32)$$

The first factor of $1/2$ is due to use of the conjugate symmetry relation (29), while the second factor of $1/2$ is due to the symmetry relation of equation (30). These factors constitute a significant reduction in the number of complex coefficients that need to be determined at second order.

A seemingly “obvious” partitioning of f_1, f_2 space is to take squares of size $W \times W$. If allowed, this would reduce the COD by the tremendous factor of $(F/W)^2$. However, it is found that the outputs of these individual frequency squares interact and contaminate each other. In addition, it is not obvious how the off-diagonal squares in f_1, f_2 , which cover different frequency ranges in f_1 and f_2 , should be fit to a *single* band of frequencies of filtered output $z(n\Delta)$.

4.3 ALTERNATIVE FORM FOR SECOND-ORDER OUTPUT $y_2(n\Delta)$

Substitution of equation (28) into equation (24), and interchange of double summation and integration, leads to the relation

$$y_2(n\Delta) = \Delta^2 \iint df_1 df_2 \exp[i 2 \pi (f_1 + f_2) n\Delta] H_2(f_1, f_2) X(f_1) X(f_2). \quad (33)$$

This is not a double Fourier transform; there is only one time variable on the right-hand side, namely, $n\Delta$.

The key observation to make at this juncture is that the *only* place that time variable $n\Delta$ appears on the right-hand side of equation (33) is with the frequency *combination* $f_1 + f_2$. If second-order Volterra output $y_2(n\Delta)$ is to have frequency content *only* in the band (f_a, f_b) , for purposes of fitting to a corresponding filtered version of $z(n\Delta)$, and if $X(f)$ is broadband, then second-order frequency-domain kernel $H_2(f_1, f_2)$ must be restricted to be nonzero only for

$$f_a < f_1 + f_2 < f_b \quad (34)$$

(and the corresponding negative frequencies). This condition allows the complex exponential in equation (33) to take on frequency variation only in the band (f_a, f_b) . The region in equation (34) is definitely *not* square in f_1, f_2 space. Rather, see the blue and green regions in figure 2.

Equation (34) describes an infinite strip at angle -45° in the f_1, f_2 plane, with perpendicular width $(f_b - f_a)/\sqrt{2} = W/\sqrt{2}$. However, the fundamental region is limited to be below the $+45^\circ$ line in the f_1, f_2 plane; see the red-bordered 90° east region in figure 2. In addition, frequency f_1 cannot exceed the limit F . The shape of this finite confined strip in the f_1, f_2 plane is similar to the shape of the state of Nevada. This is the restricted region of f_1, f_2 space in which $H_2(f_1, f_2)$ is allowed to be nonzero if $y_2(n\Delta)$ in equation (33) is to contain frequency content limited to the frequency range (f_a, f_b) .

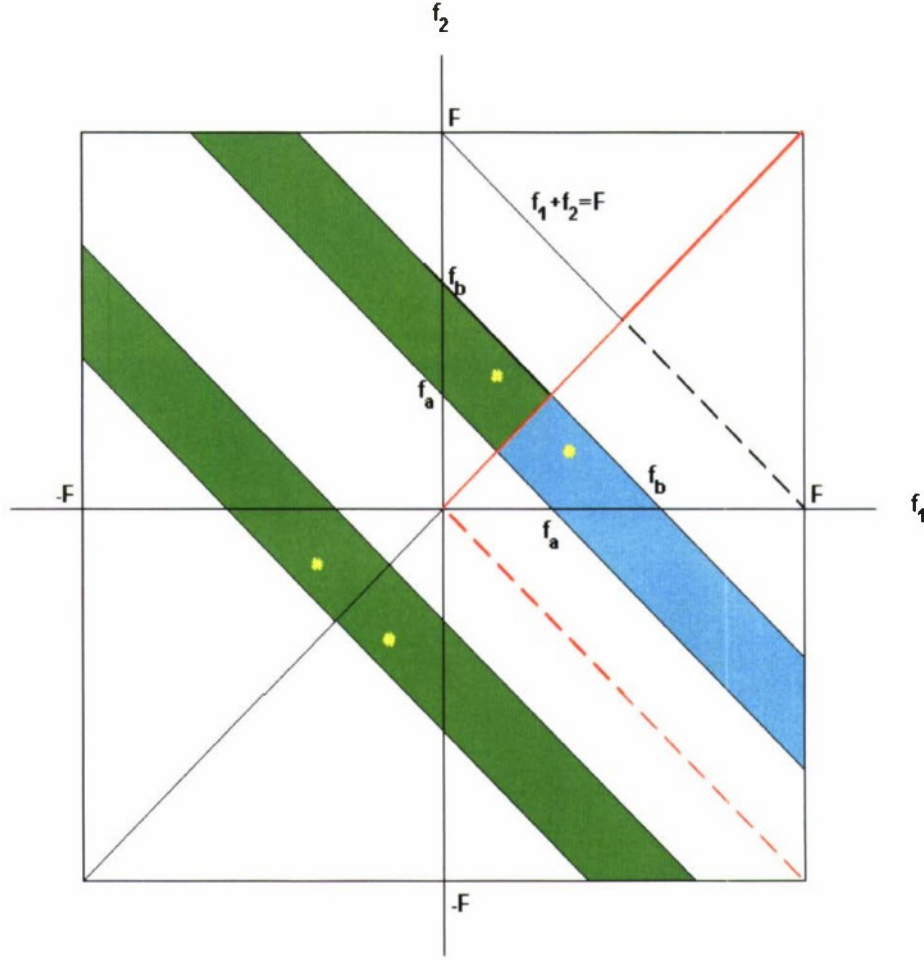


Figure 2. Diagonal Strips in f_1, f_2 Plane

Frequency f_a can start at zero. Also, it is not necessary to let frequency f_b exceed F . It is assumed that there is no frequency content in nonlinearity output $z(n\Delta)$ beyond F . Thus, the fundamental (blue) strip can slide anywhere between the dashed red and black lines. The four yellow Xs represent the locations of the symmetry points of $H_2(f_1, f_2)$ dictated by equations (29) and (30).

The length of the frequency strip (34) along the -45° line starts at value $\sqrt{2} F$ for $f_a = 0$. Therefore, the area covered by each strip is approximately $\sqrt{2} F \times W / \sqrt{2} \approx FW$, not F^2 as it would be if frequency partitioning were not employed at all. Thus, the COD has been alleviated (not eliminated) by a factor of W/F through frequency partitioning, a very worthwhile and welcome reduction in many cases. Since the frequency “area” covered by each elemental complex exponential in equation (17) is $(1/T)^2$, the number of unknown coefficients is about $FW / (1/T)^2 = TF TW$, not $(TF)^2$.

4.4 ALTERNATIVE DERIVATION OF RESULT (34)

A Fourier transformation of equation (33) results in spectrum

$$Y_2(f) = \int df_1 H_2(f_1, f - f_1) X(f_1) X(f - f_1). \quad (35)$$

In order for $Y_2(f)$ to be nonzero only in frequency band (f_a, f_b) , function $H_2(f_1, f - f_1)$ must be nonzero only for $f_a < f < f_b$. Let $f_2 = f - f_1$. Then, $f = f_1 + f_2$, from which it then follows that condition (34) must be satisfied.

4.5 BASIS FUNCTIONS IN TWO DIMENSIONS

The basis functions in one dimension (first order) were derived in equations (17) and (18), starting from an elemental complex exponential in the fundamental range. Here, in two dimensions (second order), the corresponding complex elemental starting function is

$$[a(m_1, m_2) - i b(m_1, m_2)] \exp\left(i 2 \pi \frac{m_1}{T} k_1 \Delta + i 2 \pi \frac{m_2}{T} k_2 \Delta\right), \quad (36)$$

where coefficients $\{a(m_1, m_2)\}$ and $\{b(m_1, m_2)\}$ are real, and frequencies $f_1 = m_1/T$ and $f_2 = m_2/T$ are confined to the fundamental (blue) region in the f_1, f_2 plane, as described earlier. The *total* real function for expanding second-order time-domain kernel $h_2(k_1 \Delta, k_2 \Delta)$, using the conjugate and reflective symmetries in equations (29) and (30), respectively, is then (with scaling $1/4 = 1/2 * 1/2!$)

$$\begin{aligned} & \frac{1}{4} [a(m_1, m_2) - i b(m_1, m_2)] \exp(i A) + \frac{1}{4} [a(m_1, m_2) - i b(m_1, m_2)] \exp(i B) \\ & + \text{complex conjugate terms} \end{aligned} \quad (37)$$

$$= a(m_1, m_2) \left[\frac{\cos(A) + \cos(B)}{2} \right] + b(m_1, m_2) \left[\frac{\sin(A) + \sin(B)}{2} \right],$$

where

$$A = \frac{2 \pi \Delta}{T} (m_1 k_1 + m_2 k_2), \quad B = \frac{2 \pi \Delta}{T} (m_2 k_1 + m_1 k_2). \quad (38)$$

The two basis (bracketed) functions in the second line of equation (37) are obviously real and symmetric in k_1, k_2 , upon use of equation (38), as expected and required. However, their forms are somewhat surprising and not easy to anticipate, without the construction utilized in equation (36) and the first two lines of equation (37).

It was pointed out in the sequel to equation (15) that the coefficient $b(0)$ in the first-order case was absent because the corresponding sine term in the expansion was zero for $m = 0$. A more subtle but related property holds at second order. From equations (37) and (38), a typical term in the sine basis set is

$$\frac{1}{2} \sin\left(\frac{2\pi\Delta}{T}(m_1 k_1 + m_2 k_2)\right) + \frac{1}{2} \sin\left(\frac{2\pi\Delta}{T}(m_2 k_1 + m_1 k_2)\right). \quad (39)$$

This term is obviously zero for the point $m_1 = 0, m_2 = 0$ in f_1, f_2 space. But equation (39) is also zero for $m_1 + m_2 = 0$, even though *neither* term is zero. This latter condition yields a *line* in the two-dimensional frequency plane f_1, f_2 . All of the terms corresponding to this line, $m_1 + m_2 = 0$, must be eliminated from the sine basis set in order to avoid a singular matrix in a second-order fitting process. None of the $[\cos(A) + \cos(B)]/2$ terms in equation (37) is zero.

4.6 SECOND-ORDER KERNEL EXPANSION

Combine equations (37) and (38) to form the fundamental second-order basis functions:

$$c_2(m_1, m_2; k_1, k_2) = \frac{1}{2} \cos\left[\frac{2\pi}{K}(m_1 k_1 + m_2 k_2)\right] + \frac{1}{2} \cos\left[\frac{2\pi}{K}(m_1 k_2 + m_2 k_1)\right], \quad (40)$$

$$s_2(m_1, m_2; k_1, k_2) = \frac{1}{2} \sin\left[\frac{2\pi}{K}(m_1 k_1 + m_2 k_2)\right] + \frac{1}{2} \sin\left[\frac{2\pi}{K}(m_1 k_2 + m_2 k_1)\right],$$

using $T = K\Delta$. Both of these basis functions are symmetric in k_1, k_2 , and are symmetric in m_1, m_2 . Also,

$$c_2(0, 0; k_1, k_2) \neq 0; \quad s_2(m_1, -m_1; k_1, k_2) \equiv 0 \text{ for all } k_1, k_2. \quad (41)$$

For notational convenience, define

$$c(m, k) = \cos\left(\frac{2\pi}{K} m k\right), \quad s(m, k) = \sin\left(\frac{2\pi}{K} m k\right). \quad (42)$$

Then, equation (40) can be expanded as

$$\begin{aligned}
c_2(m_1, m_2; k_1, k_2) &= \frac{1}{2} [c(m_1, k_1) c(m_2, k_2) - s(m_1, k_1) s(m_2, k_2) \\
&\quad + c(m_1, k_2) c(m_2, k_1) - s(m_1, k_2) s(m_2, k_1)], \\
s_2(m_1, m_2; k_1, k_2) &= \frac{1}{2} [s(m_1, k_1) c(m_2, k_2) + c(m_1, k_1) s(m_2, k_2) \\
&\quad + s(m_1, k_2) c(m_2, k_1) + c(m_1, k_2) s(m_2, k_1)].
\end{aligned} \tag{43}$$

Now, referring back to equation (37), the expansion for the second-order kernel is given by

$$\begin{aligned}
h_2(k_1\Delta, k_2\Delta) &= \frac{1}{2} \sum_{M_c} a(m_1, m_2) [c(m_1, k_1) c(m_2, k_2) - s(m_1, k_1) s(m_2, k_2) \\
&\quad + c(m_1, k_2) c(m_2, k_1) - s(m_1, k_2) s(m_2, k_1)] \\
&\quad + \frac{1}{2} \sum_{M_s} b(m_1, m_2) [s(m_1, k_1) c(m_2, k_2) + c(m_1, k_1) s(m_2, k_2) \\
&\quad + s(m_1, k_2) c(m_2, k_1) + c(m_1, k_2) s(m_2, k_1)],
\end{aligned} \tag{44}$$

where M_c and M_s correspond to the particular two-dimensional region of frequency space f_1, f_2 in figure 2 that is under investigation. Region M_s is smaller than M_c by virtue of the comments following equation (39). Second-order kernel $h_2(k_1\Delta, k_2\Delta)$ in equation (44) is symmetric in k_1, k_2 .

4.7 SECOND-ORDER WAVEFORM EXPANSION

As in equations (22) and (42), define sequences

$$\begin{aligned}
x_c(n, m) &= \sum_{k=0}^{K-1} x(n-k) c(m, k), \\
x_s(n, m) &= \sum_{k=0}^{K-1} x(n-k) s(m, k).
\end{aligned} \tag{45}$$

Sequence $\{x_c(n, m)\}$ is even in m , while $\{x_s(n, m)\}$ is odd in m . Also, $x_s(n, 0) = 0$ for all n . These convolution sequences can be calculated just once and stored in two $N \times M$ arrays for further processing as needed. Substitution of equation (44) into equation (24), interchange of the double summations, and the use of equation (45) yields the second-order Volterra component in the form

$$\begin{aligned}
y_2(n) = & \sum_{M_c} a(m_1, m_2) [x_c(n, m_1) x_c(n, m_2) - x_s(n, m_1) x_s(n, m_2)] \\
& + \sum_{M_s} b(m_1, m_2) [x_s(n, m_1) x_c(n, m_2) + x_c(n, m_1) x_s(n, m_2)],
\end{aligned} \tag{46}$$

where duplicate terms have been added together. Both bracketed terms are symmetric in m_1, m_2 . This last equation for $y_2(n)$ is now in a form ready for use in a least squares fit to a filtered version of nonlinearity output $z(n)$.

4.8 DISCARDING OF EDGE STRIPS AT SECOND ORDER

In an earlier first-order subsection, a method for the discarding of edge coefficients for the first-order fitting procedure was presented. That method can be generalized to second order as follows. The 1-kHz *diagonal strip* next to the bottom left edge of each 6-kHz diagonal strip in the f_1, f_2 plane must be discarded. That particular 1-kHz diagonal strip in f_1, f_2 leads to the frequency components of $y_2(n\Delta)$ at the lower edge of the (f_a, f_b) band, namely, to components in the band $(f_a, f_a + 1 \text{ kHz})$.

Similarly, the 1-kHz diagonal strip next to the upper right edge of each 6-kHz diagonal strip must be discarded. Only the interior 4-kHz diagonal strip of each 6-kHz diagonal strip is retained. When the adjacent frequency band comes under investigation for fitting purposes, overlap will again be required. But the next analysis strip will be moved over by 4 kHz, not 6 kHz. Thus, two edge *bands* are discarded at first order, but two edge *strips* are discarded at second order. Third order will likely require discarding of edge *volumes* in the three-dimensional frequency space f_1, f_2, f_3 .

4.9 TEST PROCEDURE FOR A FIRST-ORDER AND SECOND-ORDER EXAMPLE

A nonlinear system with known first-order kernel $h_{1e}(k_1\Delta)$ and known second-order kernel $h_{2e}(k_1\Delta, k_2\Delta)$ was excited with stationary random process $x(n\Delta)$. The output $z_1(n\Delta)$ of the first-order kernel h_{1e} was computed and stored. Also, the output $z_2(n\Delta)$ of the second-order kernel h_{2e} was computed and stored. Finally, the total nonlinear system output

$$z(n\Delta) = z_0 + z_1(n\Delta) + z_2(n\Delta) \tag{47}$$

was computed and stored, as was colored excitation $x(n\Delta)$. See figure 3.

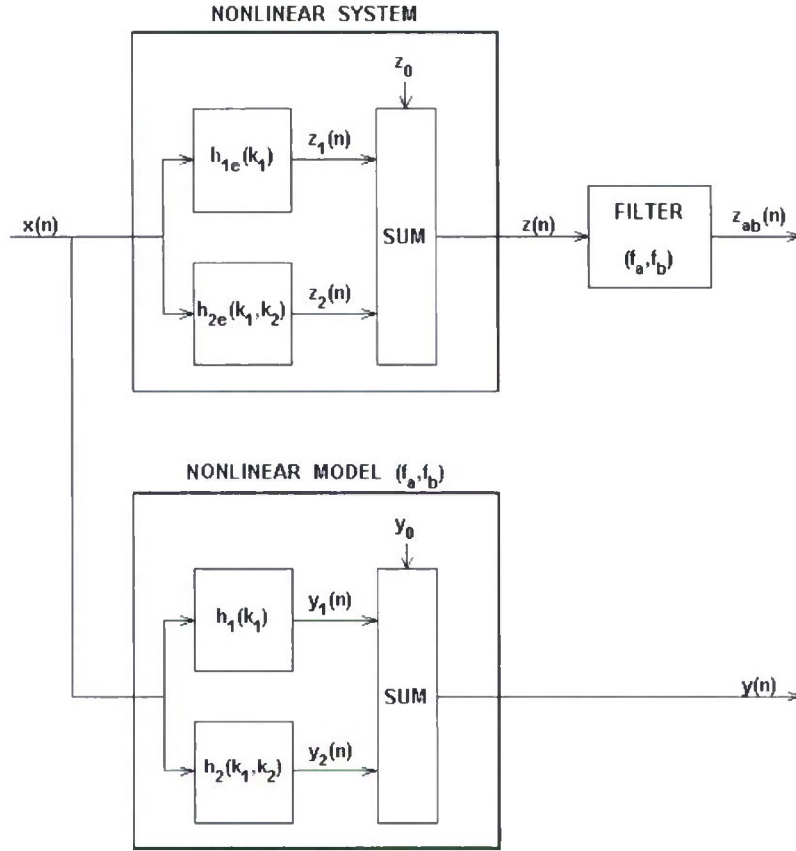


Figure 3. Test Procedure for Second Order

Working only with the recordings of excitation $x(n\Delta)$ and total nonlinear response $z(n\Delta)$, the object was to choose real coefficients $a(m)$ and $b(m)$ for first-order Volterra kernel $h_1(k_1\Delta)$, and to simultaneously choose real coefficients $a(m_1, m_2)$ and $b(m_1, m_2)$ for second-order Volterra kernel $h_2(k_1\Delta, k_2\Delta)$, such that the total Volterra model output

$$y(n\Delta) = y_0 + y_1(n\Delta) + y_2(n\Delta) \quad (48)$$

fitted the *filtered* waveform $z_{ab}(n\Delta)$ as well as possible in a least squares sense.

Five checks can be conducted on the estimates obtained from this control example. Namely, one can compare: estimate h_1 with the exact h_{1e} ; estimate h_2 with the exact h_{2e} ; estimate y_1 with the exact z_1 ; estimate y_2 with the exact z_2 ; and estimate y with the exact z . In an actual practical situation of an *unknown* nonlinear system under investigation, the only available comparison will be least squares model response y with nonlinearity response z . This comparison can also be conducted band-by-band in frequency.

4.10 INTERPRETATION AND USEFULNESS

The power in the total model output $y(n\Delta)$ relative to the power in total nonlinear output $z(n\Delta)$ measures the goodness of the fit. If only a small fraction of the $z(n\Delta)$ power is accounted for, there may be independent noise generated inside the nonlinear system that can't possibly be fit; the level of this internal noise can be determined by turning off input excitation $x(n\Delta)$ and observing the inherent power in output $z(n\Delta)$. Or the order of the Volterra (or Wiener) model may simply be inadequate and must be increased for a better fit. This would indicate that higher order nonlinearities are present in the system under investigation.

The relative levels of the powers in the total final fits $y_1(n\Delta)$ and $y_2(n\Delta)$ can be used as indicators of the relative amount of second-order nonlinearity in the system of interest. If the nonlinear power is significant in comparison with the linear power, it will behoove one to account for this nonlinearity in the system in the future.

Volterra kernel estimates $h_1(k_1\Delta)$ and $h_2(k_1\Delta, k_2\Delta)$ can now be used to *predict* the performance of the investigated nonlinear system to different future inputs, without actually having to perform those experiments.

5. SECOND-ORDER NUMERICAL RESULTS FOR THE CONTROL EXAMPLE

The following results were accomplished by frequency partitioning. Figure 4 displays the results for the first-order kernel. The exact kernel $h_{1e}(k_1) = h_{11}(k_1)$ is drawn in blue (b), and the estimated model kernel $h_1(k_1)$ is drawn in black (k). The error $e_{h1}(k_1)$ between them is drawn in red (r). The estimate is a virtual overlay of the exact kernel; there is a small region at the far right where the red error curve deviates slightly from zero. The ratio of the energy in the error sequence to the energy in the exact kernel sequence is 0.000042.

The corresponding results for the second-order kernel $h_{2e}(k_1, k_2) = h_{22}(k_1, k_2)$ are given in figures 5 and 6. The estimated model kernel in figure 6 is visually indistinguishable from the exact kernel in figure 5. The ratio of the energy in the error between the two functions to the energy in the exact kernel is 0.00023.

Figure 7 gives the results for a section of the first-order waveforms. The error waveform $e_1(n)$ between estimate $y_1(n)$ and exact waveform $z_1(n)$ is extremely small and has a relative error energy of 0.000043, the same as for the first-order kernels in figure 4.

The corresponding comparison for the second-order waveforms is presented in figure 8. The variation with time of these second-order waveforms is much more erratic and quicker than for the first-order results in figure 7. This is not unexpected for a second-order nonlinearity. The ratio of error energy to exact energy is 0.00014.

Finally, the total waveform results for estimate $y(n)$ and nonlinearity response waveform $z(n)$ are given in figure 9. The error energy ratio is 0.000042 for this case. This last plot is the only comparison that can be constructed for the practical examples where only input excitation $x(n)$ and nonlinear response $z(n)$ are available.

Two new concepts have been introduced at second order in this investigation. They are the fundamental region (31) and the special basis functions (40) in two-dimensional time space k_1, k_2 . These items are crucial in achieving a satisfactory second-order fit by means of least squares. Also, it was then necessary to eliminate any zero basis functions, and to use overlap and to discard edge estimates that were untrustworthy. The combination of all these items then led to a viable and accurate method of estimating the internal behavior of a second-order nonlinear system with memory.

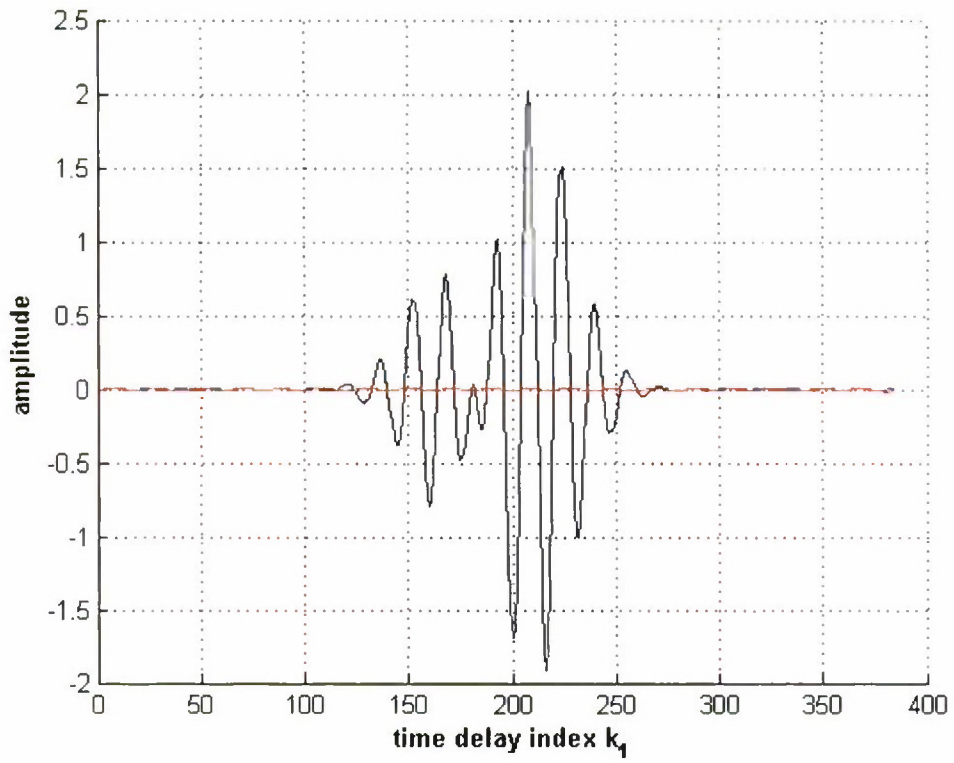


Figure 4. First-Order Kernels $h_{11}(b)$ $h_1(k)$ $e_{h1}(r)$

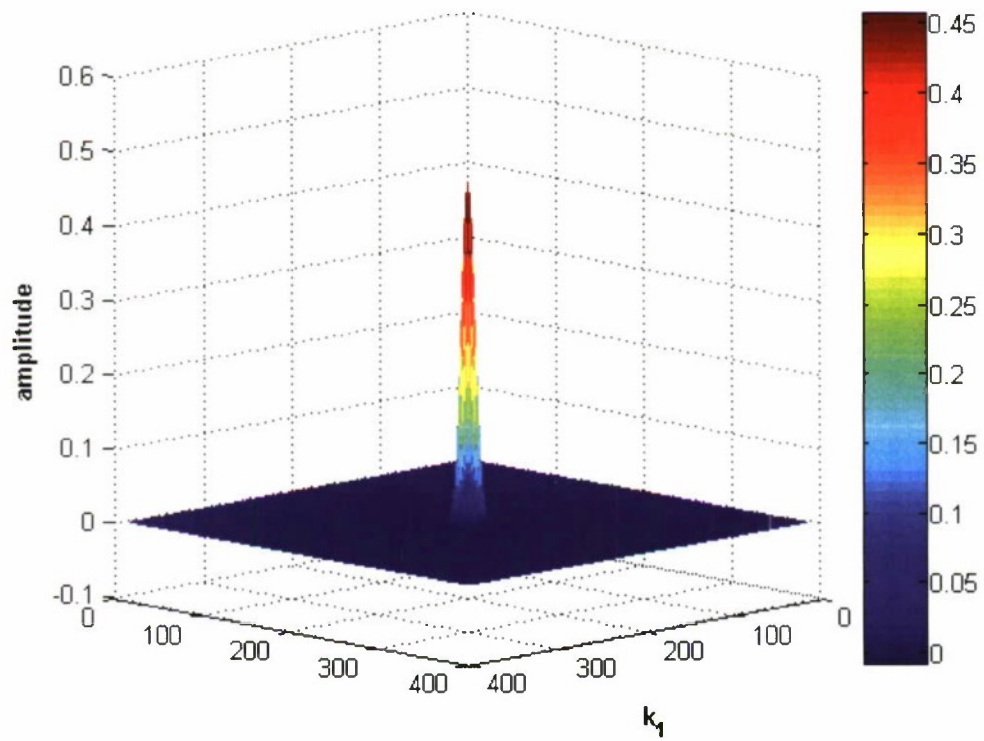


Figure 5. Exact Second-Order Kernel $h_{22}(k_1, k_2)$

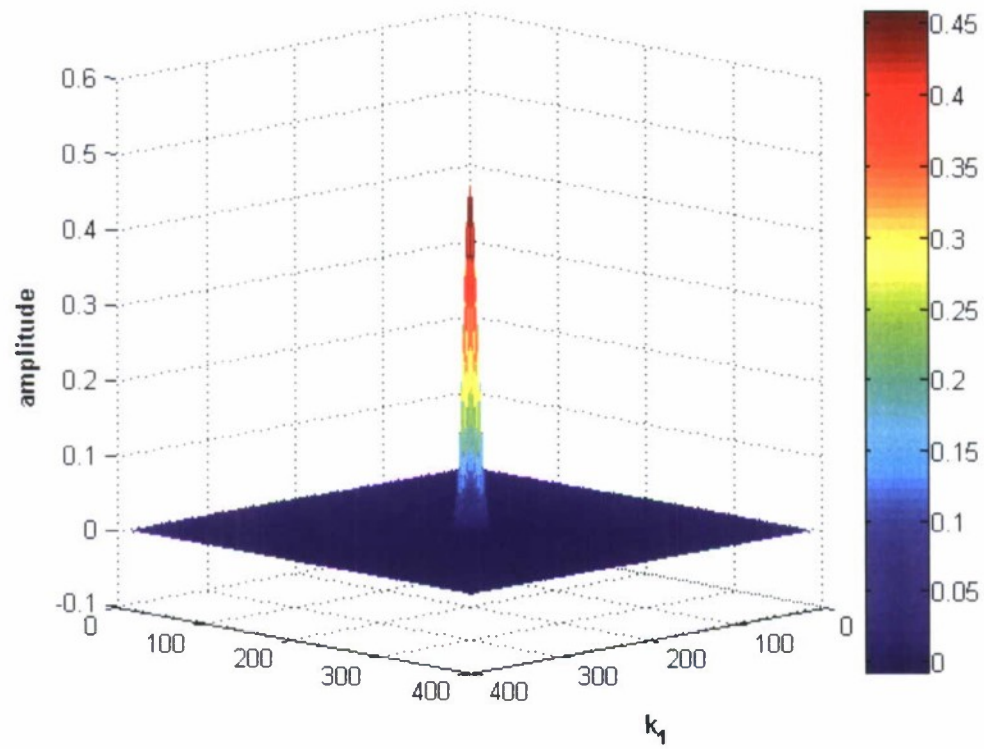


Figure 6. Model Second-Order Kernel $h_2(k_1, k_2)$

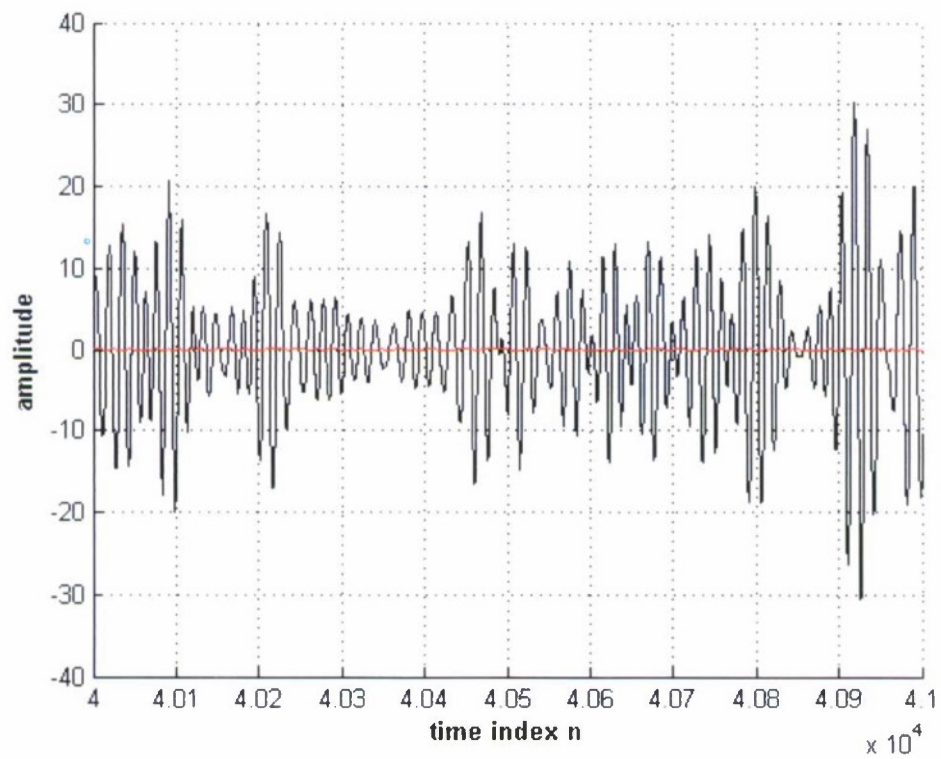


Figure 7. First-Order Waveforms $z_1(b)$ $y_1(k)$ $e_1(r)$

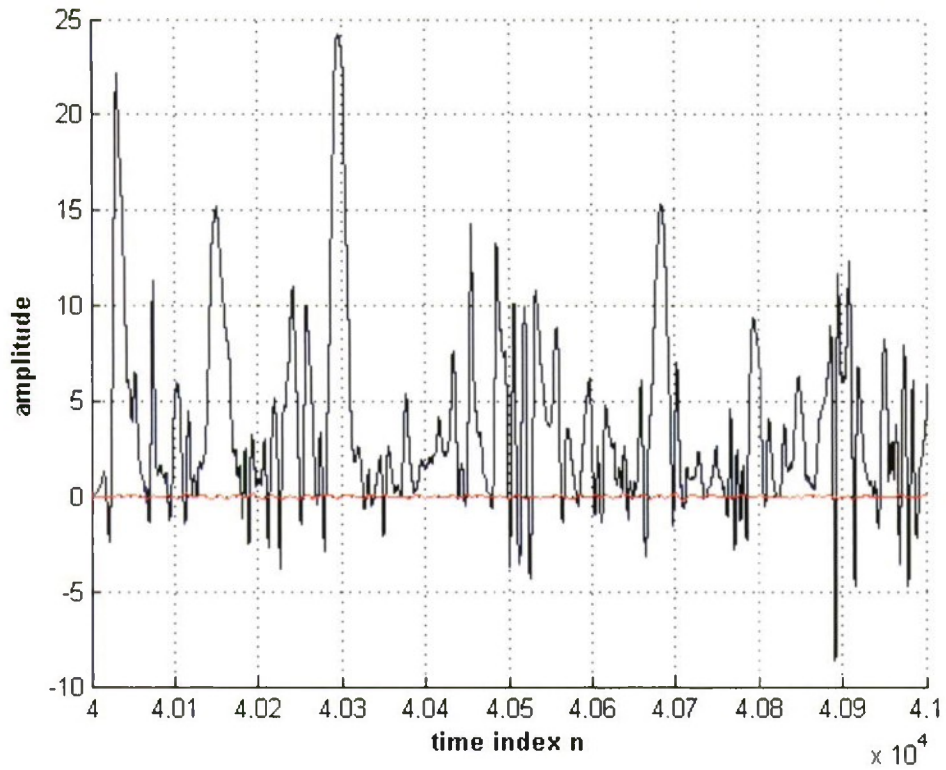


Figure 8. Second-Order Waveforms $z_2(b)$ $y_2(k)$ $e_2(r)$

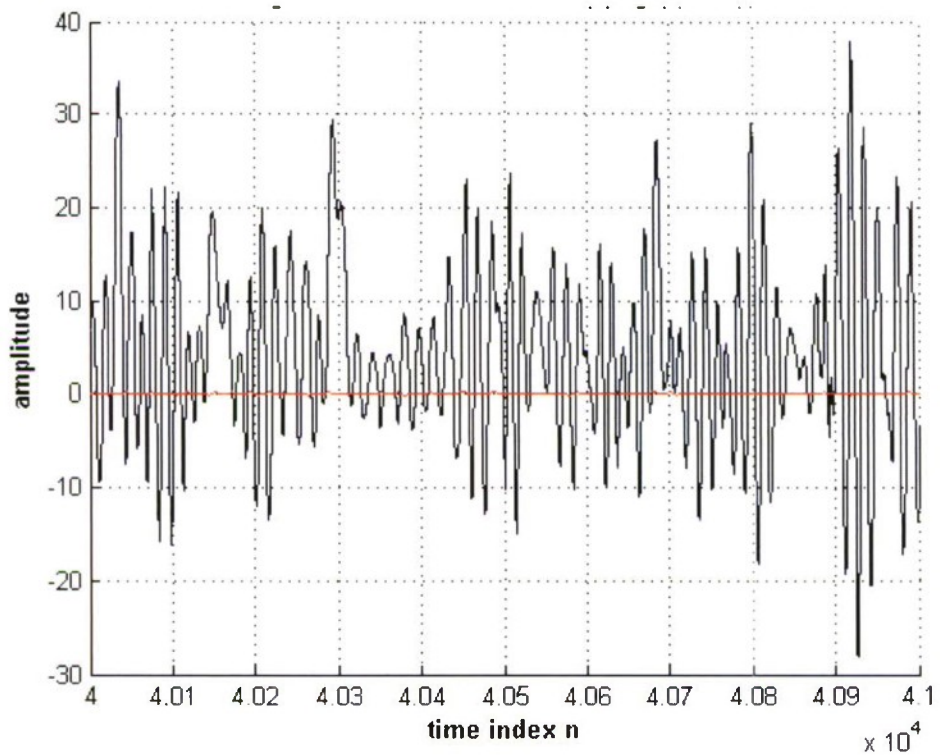


Figure 9. Waveforms $z(b)$ $y(k)$ $e(r)$

6. THIRD-ORDER CONSIDERATIONS

6.1 UNIQUENESS REQUIREMENT

The third-order Volterra component waveform was given in equation (1) as

$$y_3(n) = \sum_{k_1=0}^{K-1} \sum_{k_2=0}^{K-1} \sum_{k_3=0}^{K-1} h_3(k_1, k_2, k_3) x(n-k_1) x(n-k_2) x(n-k_3). \quad (49)$$

It must be noted immediately that there can be no unique solution for the third-order time-domain kernel $h_3(k_1, k_2, k_3)$. Only the *sum*

$$h_3(k_1, k_2, k_3) + h_3(k_3, k_1, k_2) + h_3(k_2, k_3, k_1) + h_3(k_1, k_3, k_2) + h_3(k_2, k_1, k_3) + h_3(k_3, k_2, k_1) \quad (50)$$

of six symmetrically located kernel values in k_1, k_2, k_3 space affects output $y_3(n)$. This is due to the symmetry of the product of the three x values in variables k_1, k_2, k_3 in equation (49). To enable a unique solution for the third-order time-domain kernel, it is necessary to impose some condition. The condition adopted here is to take the third-order kernel to be real and symmetric:

$$h_3(k_1, k_2, k_3) = h_3(k_3, k_1, k_2) = h_3(k_2, k_3, k_1) = h_3(k_1, k_3, k_2) = h_3(k_2, k_1, k_3) = h_3(k_3, k_2, k_1). \quad (51)$$

There is *no* loss of generality in imposing this symmetry relation on the third-order time-domain kernel $h_3(k_1, k_2, k_3)$, because the output $y_3(n)$ of equation (49) depends only on the sum in equation (50), no matter what the excitation $x(n)$ is. Also, significant advantage will be taken of this symmetry property, including a reduction in the number of unknown coefficients that must be determined, as well as avoidance of singular matrices.

6.2 PROPERTIES OF THE THIRD-ORDER FREQUENCY-DOMAIN KERNEL

The third-order frequency-domain (complex) kernel corresponding to $h_3(k_1\Delta, k_2\Delta, k_3\Delta)$ is

$$H_3(f_1, f_2, f_3) = \sum_{k_1=0}^{K-1} \sum_{k_2=0}^{K-1} \sum_{k_3=0}^{K-1} h_3(k_1\Delta, k_2\Delta, k_3\Delta) \exp(-i2\pi f_1 k_1\Delta - i2\pi f_2 k_2\Delta - i2\pi f_3 k_3\Delta). \quad (52)$$

The inverse relation is

$$h_3(k_1\Delta, k_2\Delta, k_3\Delta) = \Delta^3 \iiint df_1 df_2 df_3 H_3(f_1, f_2, f_3) \exp(i2\pi f_1 k_1\Delta + i2\pi f_2 k_2\Delta + i2\pi f_3 k_3\Delta). \quad (53)$$

Let the time duration of $h_3(k_1\Delta, k_2\Delta, k_3\Delta)$ in each time dimension be T seconds. And let the frequency extent of $H_3(f_1, f_2, f_3)$ in each frequency dimension be $(-F, F)$ Hz. These limits are identical to those used for the first- and second-order situations. Then, using symmetry, approximately $(2TF)^3/6$ samples are required to completely characterize the third-order time-domain kernel $h_3(k_1\Delta, k_2\Delta, k_3\Delta)$. This rapid increase in the number of required coefficients for characterization, in progressing from first order to second order to third order, is the COD.

Since $h_3(k_1\Delta, k_2\Delta, k_3\Delta)$ is forced to be real, it follows immediately from equation (52) that

$$H_3(-f_1, -f_2, -f_3) = H_3(f_1, f_2, f_3)^*. \quad (54)$$

And, from equations (51) and (52), there follows

$$\begin{aligned} H_3(f_1, f_2, f_3) &= H_3(f_3, f_1, f_2) = H_3(f_2, f_3, f_1) \\ &= H_3(f_1, f_3, f_2) = H_3(f_2, f_1, f_3) = H_3(f_3, f_2, f_1). \end{aligned} \quad (55)$$

Thus, the complex third-order frequency-domain kernel $H_3(f_1, f_2, f_3)$ satisfies a conjugate symmetry relation through the origin of three-dimensional f_1, f_2, f_3 space, as well as numerous symmetry relations.

6.3 ALTERNATIVE FORM FOR THIRD-ORDER OUTPUT $y_3(n\Delta)$

Substitution of equation (53) into equation (49), and interchange of triple summation and integration, leads to the relation

$$y_3(n\Delta) = \Delta^3 \iiint df_1 df_2 df_3 \exp[i 2\pi (f_1 + f_2 + f_3)n\Delta] H_3(f_1, f_2, f_3) X(f_1) X(f_2) X(f_3). \quad (56)$$

This is not a triple Fourier transform; there is only one time variable on the right-hand side, namely, $n\Delta$.

The key observation to make at this juncture is that the *only* place that time variable $n\Delta$ appears on the right-hand side of equation (56) is with the frequency *combination* $f_1 + f_2 + f_3$. If third-order Volterra output $y_3(n\Delta)$ is to have frequency content *only* in the band (f_a, f_b) , for purposes of fitting to a corresponding filtered version of $z(n\Delta)$, and if $X(f)$ is broadband, then frequency-domain kernel $H_3(f_1, f_2, f_3)$ must be restricted to be nonzero only for

$$f_a < f_1 + f_2 + f_3 < f_b \quad (57)$$

(and the corresponding negative frequencies). This condition allows the complex exponential in equation (56) to take on frequency variations only in the band (f_a, f_b) . The lower bound in equation (57) describes a plane in three dimensions, while the upper bound describes a parallel plane. The in-between region in equation (57) is an infinite tilted slab in three dimensions.

The restricted frequency region at second order was given in equation (31). That result can be extended to third order by observing that only one of the six equal third-order frequency-domain values in equation (55) needs to be specified. In particular, one selects as the primary case that which satisfies the rule

$$|f_3| \leq |f_2| \leq |f_1|. \quad (58)$$

This limited three-dimensional frequency region is further reduced by application of equation (54), leading to the final restricted region

$$|f_3| \leq |f_2| \leq f_1, \quad 0 \leq f_1 \leq F. \quad (59)$$

If frequency f_1 is held fixed at a positive value, this allowed region in the f_2, f_3 plane consists of two 90° sectors, namely, the east and west sectors. Thus, the shape of the restricted *volume* in f_1, f_2, f_3 space is a pair of triangular pyramids with one common edge.

Advantage must always be taken of these symmetry properties because they afford a significant reduction in the number of unknowns that must be determined (estimated). Only the values of $H_3(f_1, f_2, f_3)$ in the limited region (59) need to be specified.

The volume of the complete frequency space f_1, f_2, f_3 , of extent $(-F, F)$ in each dimension, is $(2F)^3$. On the other hand, the volume covered by equation (59) is only

$$\int_0^F df_1 \int_{-f_1}^{f_1} df_2 \int_{-|f_2|}^{|f_2|} df_3 = \int_0^F df_1 \int_{-f_1}^{f_1} df_2 \, 2|f_2| = \int_0^F df_1 \, 2f_1^2 = \frac{2}{3} F^3 = \frac{1}{2} \frac{1}{6} (2F)^3. \quad (60)$$

The factor of $1/2$ is due to use of the conjugate symmetry relation (54), while the factor of $1/6$ is due to the symmetry relations of equation (55). These factors constitute a significant reduction in the number of complex coefficients that need to be determined at third order.

6.4 BASIS FUNCTIONS IN THREE DIMENSIONS

Corresponding to the two-dimensional development in equation (36), the complex elemental starting function in three dimensions (third order) is

$$[a(m_1, m_2, m_3) - i b(m_1, m_2, m_3)] \exp\left(i 2\pi \frac{m_1}{T} k_1 \Delta + i 2\pi \frac{m_2}{T} k_2 \Delta + i 2\pi \frac{m_3}{T} k_3 \Delta\right), \quad (61)$$

where coefficients $\{a(m_1, m_2, m_3)\}$ and $\{b(m_1, m_2, m_3)\}$ are real, and frequencies

$$f_1 = m_1/T, \quad f_2 = m_2/T, \quad f_3 = m_3/T \quad (62)$$

are confined to the region in the f_1, f_2, f_3 plane described in equation (59). Let

$$\alpha = 2\pi\Delta/T = 2\pi/K, \quad c = a(m_1, m_2, m_3) - i b(m_1, m_2, m_3). \quad (63)$$

Then, using equation (61), the *total* elemental real function for expanding third-order time-domain kernel $h_3(k_1\Delta, k_2\Delta, k_3\Delta)$, using the conjugate and reflective symmetries in equations (54) and (55), respectively, is, after expansion, simplification, and scaling by $1/12 = 1/2 * 1/3!$,

$$\begin{aligned} & \frac{1}{12} \{c \exp[i\alpha(m_1k_1 + m_2k_2 + m_3k_3)] + c \exp[i\alpha(m_3k_1 + m_1k_2 + m_2k_3)] \\ & + c \exp[i\alpha(m_2k_1 + m_3k_2 + m_1k_3)] + c \exp[i\alpha(m_1k_1 + m_3k_2 + m_2k_3)] \\ & + c \exp[i\alpha(m_2k_1 + m_1k_2 + m_3k_3)] + c \exp[i\alpha(m_3k_1 + m_2k_2 + m_1k_3)]\} \\ & + \text{complex conjugate terms} \\ & = a(m_1, m_2, m_3) v_a + b(m_1, m_2, m_3) v_b, \end{aligned} \quad (64)$$

where the real basis functions are

$$\begin{aligned} v_a &= \frac{1}{6} \{ \cos[\alpha(m_1k_1 + m_2k_2 + m_3k_3)] + \cos[\alpha(m_3k_1 + m_1k_2 + m_2k_3)] \\ & + \cos[\alpha(m_2k_1 + m_3k_2 + m_1k_3)] + \cos[\alpha(m_1k_1 + m_3k_2 + m_2k_3)] \\ & + \cos[\alpha(m_2k_1 + m_1k_2 + m_3k_3)] + \cos[\alpha(m_3k_1 + m_2k_2 + m_1k_3)] \}, \\ v_b &= \frac{1}{6} \{ \sin[\alpha(m_1k_1 + m_2k_2 + m_3k_3)] + \sin[\alpha(m_3k_1 + m_1k_2 + m_2k_3)] \\ & + \sin[\alpha(m_2k_1 + m_3k_2 + m_1k_3)] + \sin[\alpha(m_1k_1 + m_3k_2 + m_2k_3)] \\ & + \sin[\alpha(m_2k_1 + m_1k_2 + m_3k_3)] + \sin[\alpha(m_3k_1 + m_2k_2 + m_1k_3)] \}. \end{aligned} \quad (65)$$

To minimize computer storage requirements, both functions in equation (65) must be expanded and expressed in terms of the quantities $c(m, k)$ and $s(m, k)$ defined in equation (42). The end results are

$$\begin{aligned} v_a &= \frac{1}{6} [c(m_1, k_1) t_1 - s(m_1, k_1) t_2 + c(m_1, k_2) t_3 - s(m_1, k_2) t_4 + c(m_1, k_3) t_5 - s(m_1, k_3) t_6], \\ v_b &= \frac{1}{6} [s(m_1, k_1) t_1 + c(m_1, k_1) t_2 + s(m_1, k_2) t_3 + c(m_1, k_2) t_4 + s(m_1, k_3) t_5 + c(m_1, k_3) t_6], \end{aligned} \quad (66)$$

where

$$\begin{aligned}
t_1 &= c(m_2, k_2) c(m_3, k_3) - s(m_2, k_2) s(m_3, k_3) + c(m_2, k_3) c(m_3, k_2) - s(m_2, k_3) s(m_3, k_2), \\
t_2 &= s(m_2, k_2) c(m_3, k_3) + c(m_2, k_2) s(m_3, k_3) + s(m_2, k_3) c(m_3, k_2) + c(m_2, k_3) s(m_3, k_2), \\
t_3 &= c(m_2, k_3) c(m_3, k_1) - s(m_2, k_3) s(m_3, k_1) + c(m_2, k_1) c(m_3, k_3) - s(m_2, k_1) s(m_3, k_3), \\
t_4 &= s(m_2, k_3) c(m_3, k_1) + c(m_2, k_3) s(m_3, k_1) + s(m_2, k_1) c(m_3, k_3) + c(m_2, k_1) s(m_3, k_3), \\
t_5 &= c(m_2, k_1) c(m_3, k_2) - s(m_2, k_1) s(m_3, k_2) + c(m_2, k_2) c(m_3, k_1) - s(m_2, k_2) s(m_3, k_1), \\
t_6 &= s(m_2, k_1) c(m_3, k_2) + c(m_2, k_1) s(m_3, k_2) + s(m_2, k_2) c(m_3, k_1) + c(m_2, k_2) s(m_3, k_1).
\end{aligned} \tag{67}$$

6.5 THIRD-ORDER KERNEL EXPANSION

Thus, from equation (64), the expansion for the third-order kernel is given by

$$h_3(k_1, k_2, k_3) = \sum_{M_c} a(m_1, m_2, m_3) v_a + \sum_{M_s} b(m_1, m_2, m_3) v_b, \tag{68}$$

where v_a and v_b are given by equations (66) and (67). Region M_c is constituted by those m_1, m_2, m_3 values in equation (62) that satisfy the conditions in equation (59). Region M_s is identical to M_c except that the points $m_1 = 0, m_2 + m_3 = 0$ and $m_2 = 0, m_1 + m_3 = 0$ and $m_3 = 0, m_1 + m_2 = 0$ must be discarded because these particular basis functions v_b in equation (65) are zero for all k_1, k_2, k_3 .

6.6 THIRD-ORDER WAVEFORM EXPANSION

The third-order Volterra component waveform is now given by substitution of equation (68) into equation (49). After expansion, simplification, and use of definitions (42), the result is

$$\begin{aligned}
y_3(n) &= \sum_{M_c} a(m_1, m_2, m_3) [x_c(n, m_1) C(n, m_2, m_3) - x_s(n, m_1) S(n, m_2, m_3)] \\
&\quad + \sum_{M_s} b(m_1, m_2, m_3) [x_s(n, m_1) C(n, m_2, m_3) + x_c(n, m_1) S(n, m_2, m_3)],
\end{aligned} \tag{69}$$

where sets of six duplicate terms have been added together, and

$$\begin{aligned}
C(n, m_2, m_3) &= x_c(n, m_2) x_c(n, m_3) - x_s(n, m_2) x_s(n, m_3), \\
S(n, m_2, m_3) &= x_s(n, m_2) x_c(n, m_3) + x_c(n, m_2) x_s(n, m_3).
\end{aligned} \tag{70}$$

Both bracketed terms in equation (69) are symmetric in m_1, m_2, m_3 and involve only the known stored arrays x_c and x_s . Equation (69) is now in a form ready for use in a least squares fit to a filtered version of nonlinearity output $z(n)$.

7. APPLICATION TO A CICADA SONG

7.1 WAVEFORMS AND SPECTRA

A time section of a cicada measurement (number S14) is displayed in figures 10 and 11. The data record is about 4 seconds long and the sampling frequency is $f_s = 96$ kHz. The plots in figures 10 and 11 are of laser measurements L_1 and L_2 of the tymbal movements on both sides of the cicada's body. Figure 12 pertains to a microphone M located a few feet from the insect. The data segment to be analyzed here stretches from time index $n = 2.5e5$ to $n = 3.5e5$, which corresponds to approximately 1 second of data. From observation of the microphone data in figure 12, it is obvious that this is a noisy measurement. For example, although the two lasers are giving zero readings for $n = 0.4e5$ to $n = 1.3e5$, the microphone output is distinctly nonzero and noisy during this time interval. Therefore, it is expected that this noise continues during the 1-second analysis interval indicated above, and will constitute a portion of output data $z(n)$ that *cannot* be fitted by a Volterra expansion. The laser measurements are considered as two *inputs*, $x_1(n)$ and $x_2(n)$, to a nonlinear system with memory (the cicada body), while the microphone measurement is considered as the *output* $z(n)$ of that same nonlinear system.

A blown-up section of duration 0.05 second of laser L_2 is presented in figure 13. Although very noisy and jagged, there is obviously a "periodic behavior" of approximately 0.01 second. A sample time segment of 0.01 second is displayed in figure 14. The first-order probability density of the complete laser recording is approximately exponential, not Gaussian. Also, some very-high-frequency components are present in the input data.

The spectra of lasers L_1 and L_2 are plotted in figures 15 and 16. The periodic low-frequency component shows up as a narrowband component centered at 96 Hz, as expected. Both spectra have a great deal of low-frequency content and are fairly flat almost all the way out to the Nyquist frequency of 48 kHz. It appears that the cicada is capable of generating some very-high-frequency components, all the way up to 50 kHz or so. If that is the case, the 96-kHz sampling rate may not have been large enough to avoid some aliasing. Recall that the laser time recordings themselves in figures 10 and 11 are very clean; that is, there is very little noise in these laser recordings. Thus, one has a non-Gaussian and non-white process to use for excitation $x(n)$. This situation is *not* what one would choose for an excitation if the input were under one's control.

The spectrum for the microphone recording is given in figure 17. There are some pronounced humps at 6 kHz and 8 kHz that bear investigation. There is a strong (unexplained) very narrowband component centered at 39.472 kHz. And, again, the spectrum remains flat almost out to the Nyquist frequency of 48 kHz.

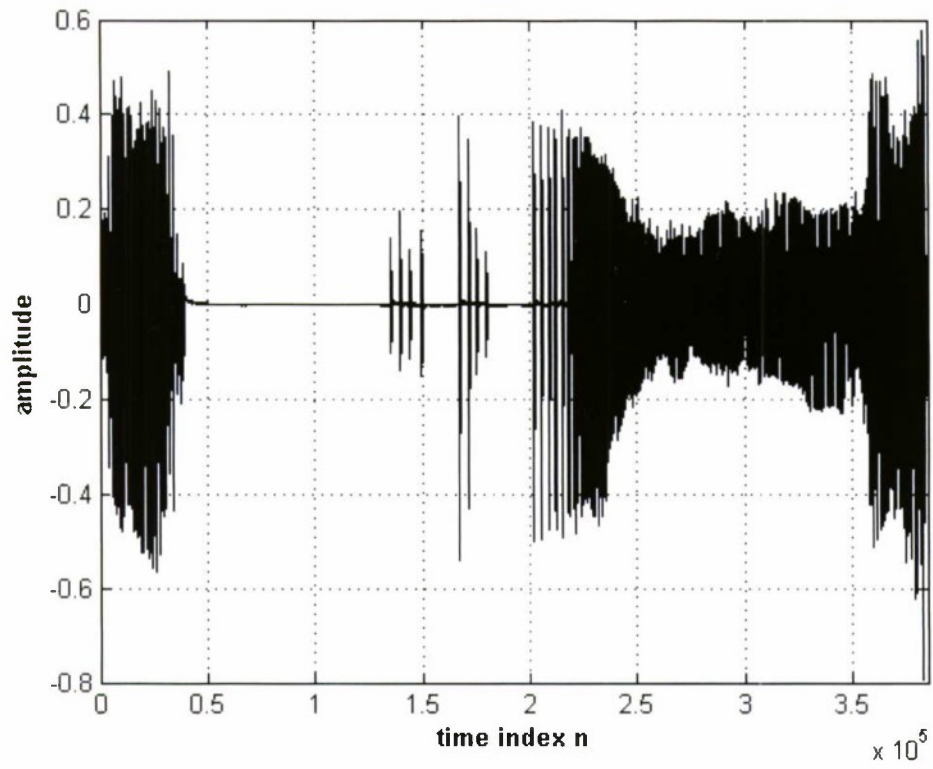


Figure 10. Waveform L_1

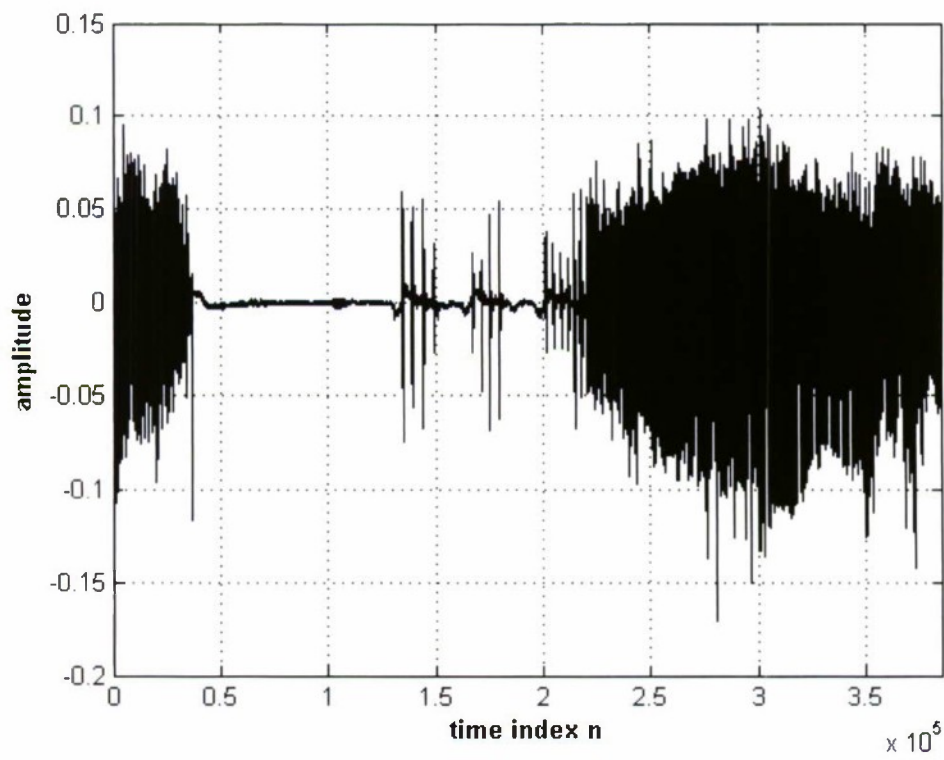


Figure 11. Waveform L_2

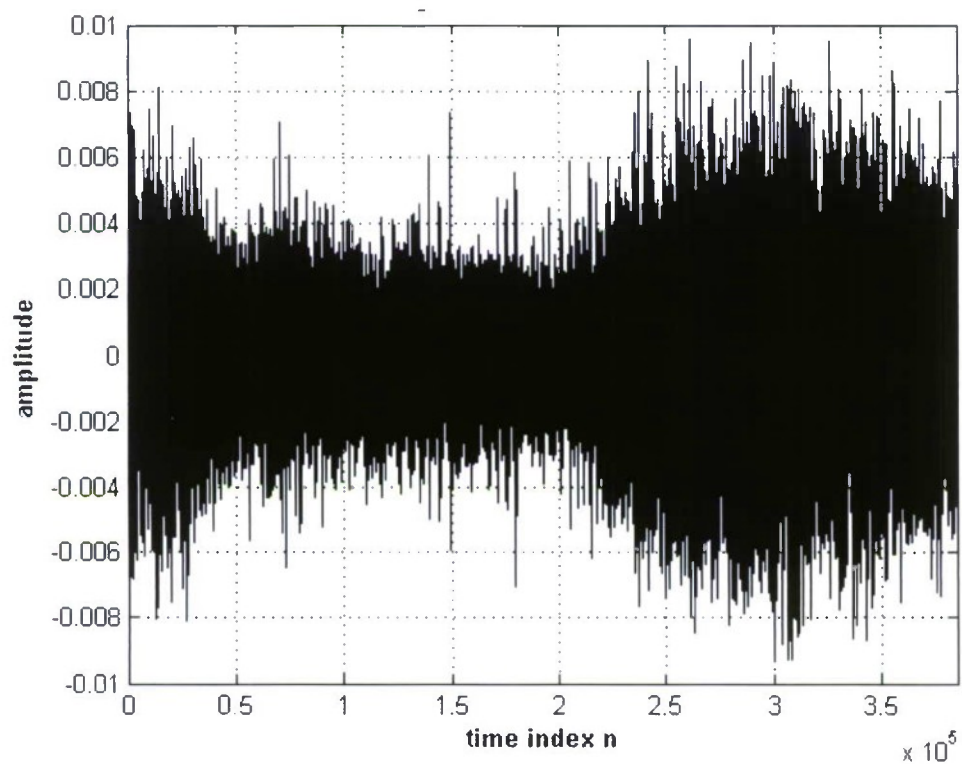


Figure 12. Waveform M

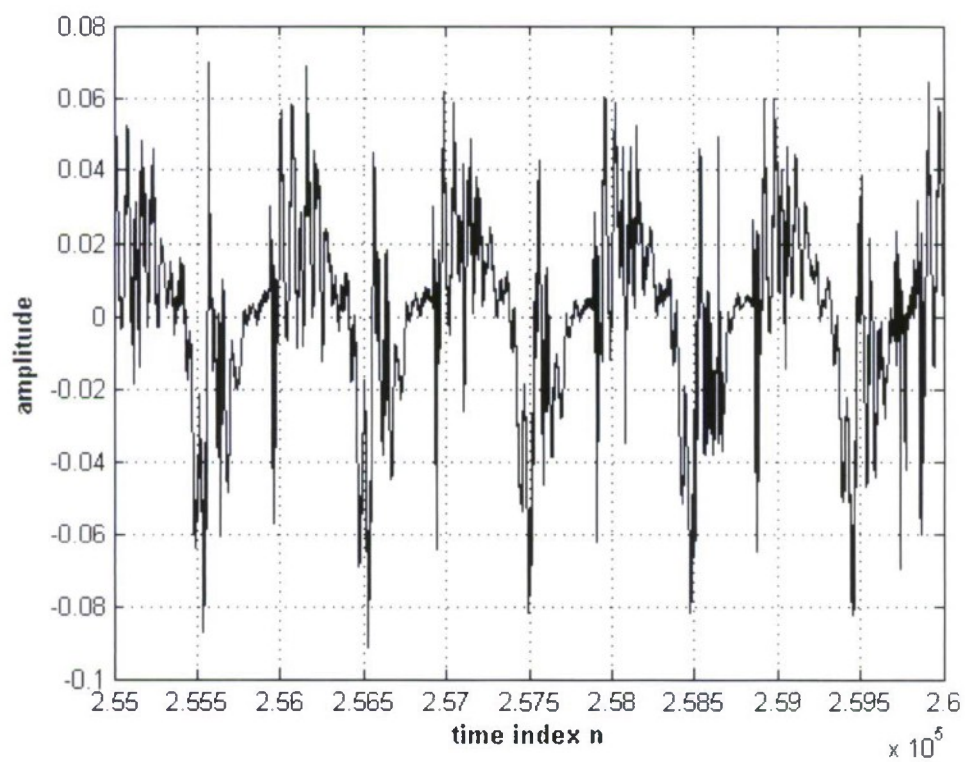


Figure 13. Waveform L_2 (0.05-Second Segment)

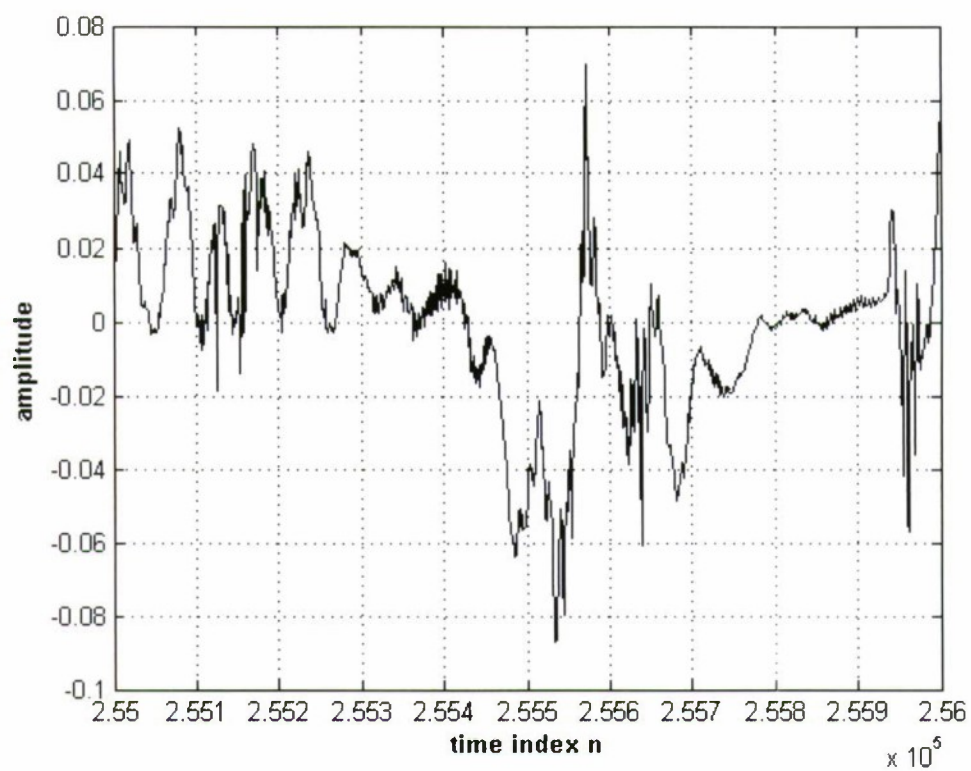


Figure 14. Waveform L_2 (0.01-Second Segment)

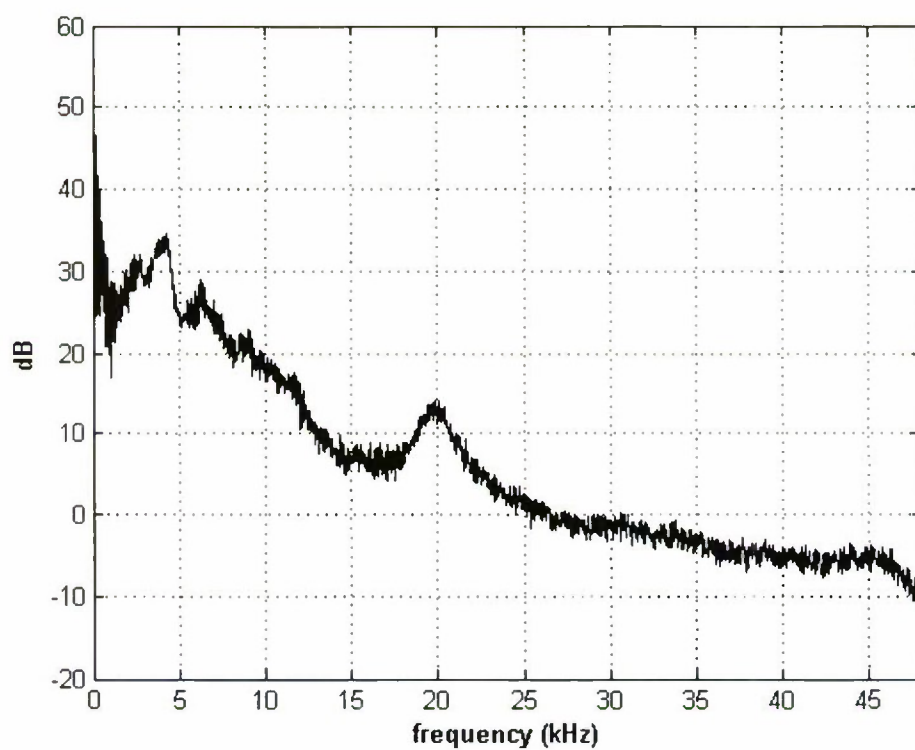


Figure 15. Spectrum of Waveform L_1

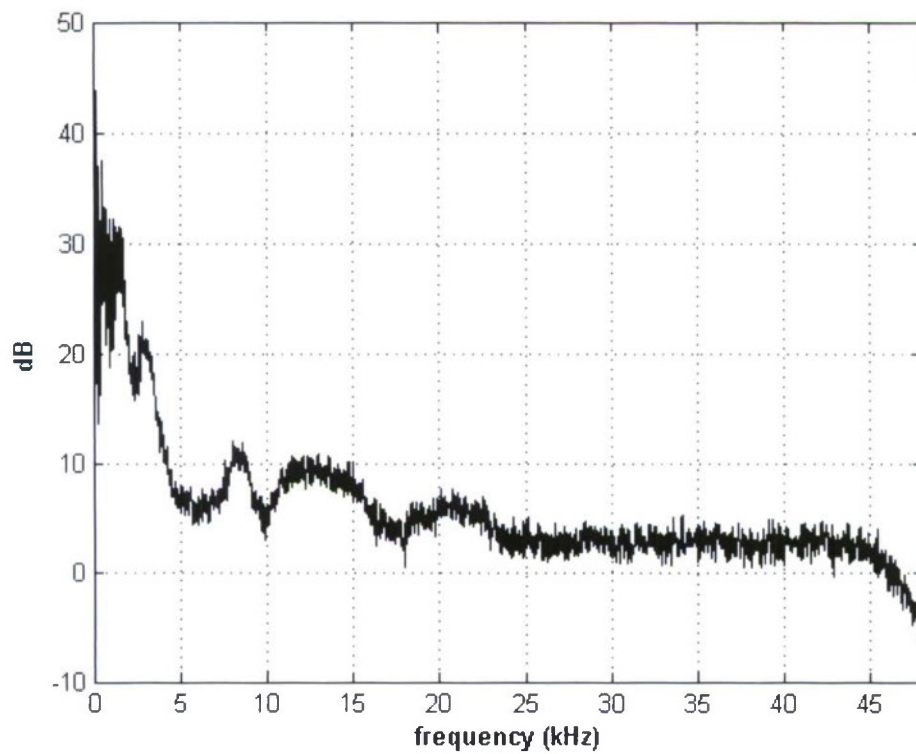


Figure 16. Spectrum of Waveform L_2

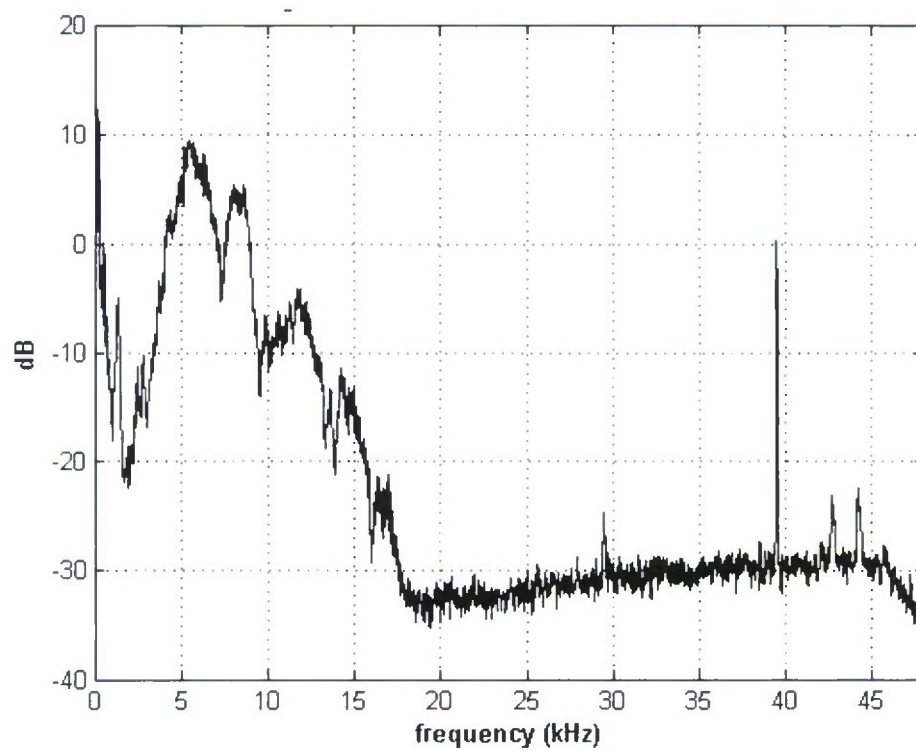


Figure 17. Spectrum of Waveform M

7.2 FIRST-ORDER FIT

A combined least squares first- and second-order fit was conducted using laser L_2 as excitation $x(n)$ and microphone M as nonlinearity response $z(n)$. The optimum first-order kernel estimate h_1 is displayed in figure 18. The memory length K was taken as 336, giving a memory duration of 3.5 ms to both the first- and second-order kernels. This duration is apparently long enough to encompass the significant contributions of these kernels, because h_1 has decayed by both ends of the fitting interval.

The corresponding estimate of the first-order frequency-domain kernel H_1 is shown in figure 19. The fitting procedure was limited to 32 kHz; the oscillations beyond this frequency are irrelevant side lobes. The major first-order response occurs between frequencies 4 kHz and 9 kHz. Thus, the high-frequency content observed in response $z(n)$ in figure 17 is not due to a linear effect on input $x(n)$, but is likely due to some nonlinear, higher order effect within the cicada's body itself.

7.3 SECOND-ORDER IDEALIZATIONS

To more easily understand some of the second-order results to be presented for this cicada investigation, it is worthwhile to first consider some idealizations of the second-order kernels, both in the two-dimensional time-delay domain τ_1, τ_2 and in the two-dimensional frequency domain f_1, f_2 . The examples will be taken in the continuous time-delay domain for simplicity; these cases have their direct counterparts in the discrete time-delay domain. The first case is

$$\begin{aligned} h_2(\tau_1, \tau_2) &= w(\tau_1) \delta(\tau_1 - \tau_2), \\ H_2(f_1, f_2) &= W(f_1 + f_2), \end{aligned} \tag{71}$$

where w is a broad function, and W is its Fourier transform. This second-order time-domain kernel is a line-impulse function along the 45° line in τ_1, τ_2 , with a slowly varying envelope w . The second-order frequency-domain kernel is a narrow mountain ridge centered along the -45° line in its domain. The height of the ridge is constant along any -45° line. Both functions in equation (71) are obviously symmetric in their respective domains. The output time waveform corresponding to equation (71) is

$$y_2(t) = \int d\tau w(\tau) x^2(t - \tau). \tag{72}$$

There is memory involved in this squaring device, but there are no "cross-terms" in which two delayed values of x are multiplied together, such as in equations (1) and (24).

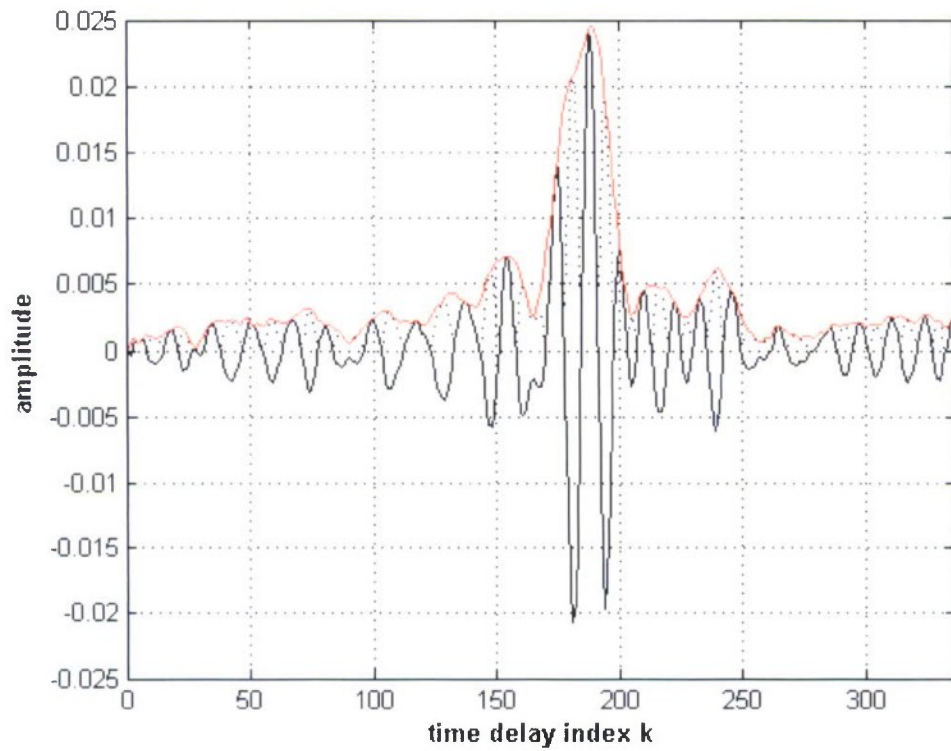


Figure 18. First-Order Kernel $h_1(k)$, $|h_1(k)|$ (:), and Envelope (r)

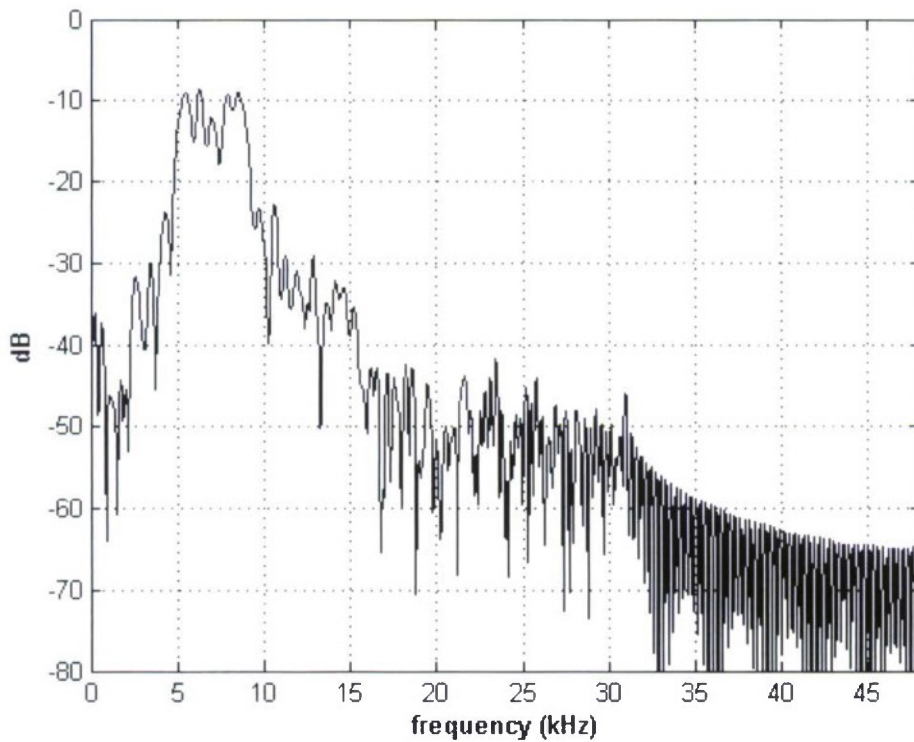


Figure 19. First-Order Kernel H_1

The second case is

$$h_2(\tau_1, \tau_2) = w\left(\frac{\tau_1 + \tau_2}{2}\right) s(\tau_1 - \tau_2), \quad (73)$$

$$H_2(f_1, f_2) = W(f_1 + f_2) S\left(\frac{f_1 - f_2}{2}\right).$$

The function s is a real, even, narrow function, in which case its Fourier transform S is a real, even, broad function. Function w is arbitrary but generally broad. Time-domain kernel h_2 has a ridge concentrated near the 45° line but has a slowly varying amplitude governed by w . Frequency-domain kernel H_2 is concentrated near the -45° line and has a slowly varying amplitude governed by S . The time-domain waveform corresponding to kernel (73) is

$$y_2(t) = \iint d\tau_1 d\tau_2 w\left(\frac{\tau_1 + \tau_2}{2}\right) s(\tau_1 - \tau_2) x(t - \tau_1) x(t - \tau_2), \quad (74)$$

which has limited-extent cross terms, governed by the duration of s .

The third case is given by

$$h_2(\tau_1, \tau_2) = 2 \cos[2\pi f_L (\tau_1 + \tau_2) + \beta] w\left(\frac{\tau_1 + \tau_2}{2}\right) s(\tau_1 - \tau_2), \quad (75)$$

$$H_2(f_1, f_2) = [\exp(i\beta) W(f_1 + f_2 - f_L) + \exp(-i\beta) W(f_1 + f_2 + f_L)] S\left(\frac{f_1 - f_2}{2}\right).$$

Frequency f_L causes the ridge of W to move up and down in the f_1, f_2 plane by the value of f_L and thereby creates two ridges parallel to the -45° line. If $s(0) = 0$, then the time-domain kernel $h_2(\tau_1, \tau_2)$ is zero for $\tau_2 = \tau_1$, which is the central $+45^\circ$ line. Combinations of different functions w and s , along with different parameter values for f_L and β lead to a wide variety of possibilities for the two second-order kernels h_2 and H_2 .

The estimate of the second-order time-domain kernel h_2 for laser L_2 of cicada song S14 is displayed in figures 20 and 21. Since this kernel has a very-high-frequency behavior, it oscillates considerably in τ_1 and τ_2 . Accordingly, it has been rectified and smoothed so that it is easier to observe where its major contributions occur. Figure 20 is for an elevation observation angle of 35° , while figure 21 is a top (90° elevation) view. There is a pronounced contribution along two symmetrically displaced lines at $\pm 45^\circ$ and a deep null along the main $+45^\circ$ line. These behaviors are consistent with those anticipated by equations (71), (73), and (75). The locations of the main energy are near the lines where $|k_2 - k_1| = 220$. At the sampling frequency of 96 kHz, this corresponds to a time delay of 2.3 ms in the cross terms of $y_2(n)$.

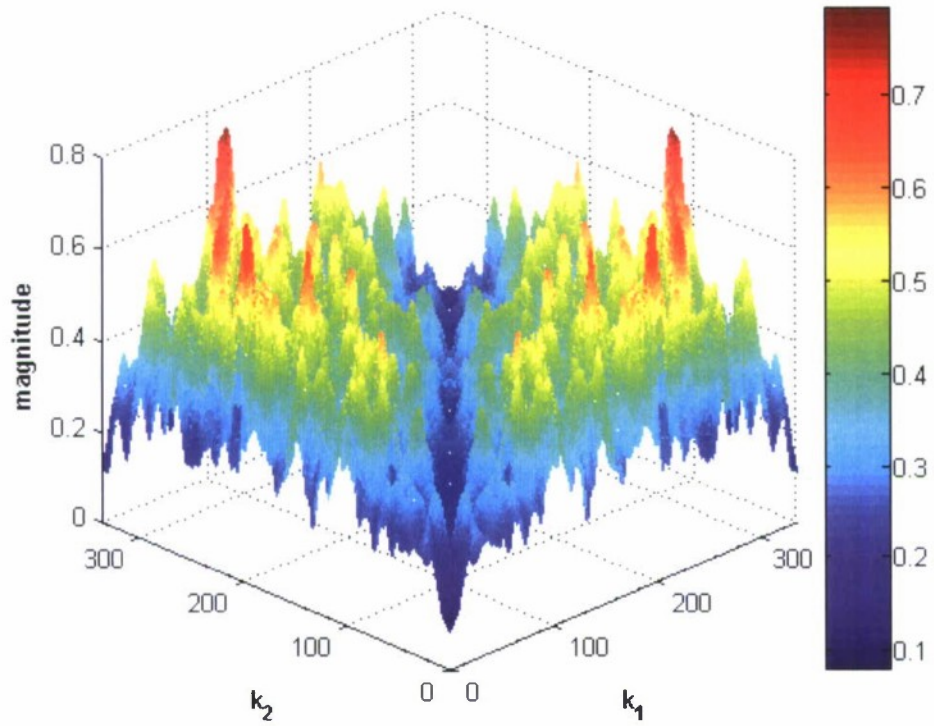


Figure 20. Smoothed Second-Order Kernel $|h_2|$ (Observation Elevation Angle 35°)

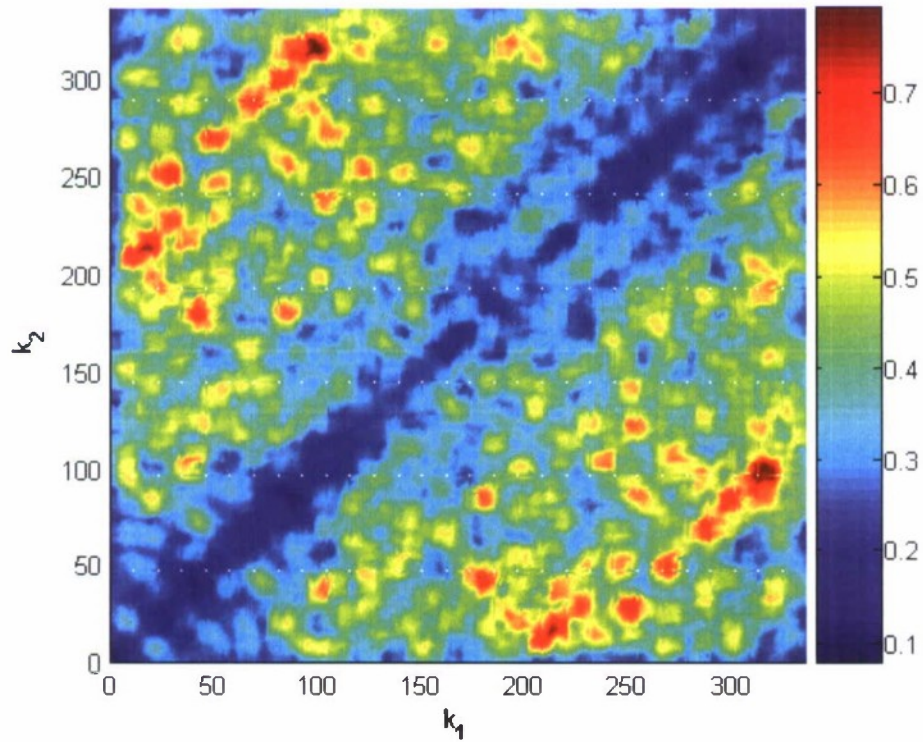


Figure 21. Smoothed Second-Order Kernel $|h_2|$ (Observation Elevation Angle 90°)

The estimate of the second-order frequency-domain kernel H_2 for laser L_2 is displayed in figures 22 and 23, the former for an observation elevation angle of 50° and the latter for an elevation angle of 90° (top view). The strong response (red line in figure 23) centered at $f_1 + f_2 = 0$ corresponds to a strong, low-frequency component in second-order model response $y_2(n)$, while the two strong spread lines centered at $f_1 + f_2 = \pm 6$ kHz correspond to the spectral hump for microphone waveform $z(n)$ in figure 17. It can be seen from figure 23 that the second-order frequency-domain kernel $H_2(f_1, f_2)$ responds very well all the way up to 32 kHz, which was the highest frequency fitted. But that large response only takes place in the second and fourth quadrants of f_1, f_2 space, which is the region where difference frequencies are created. There is very little response in the first and third quadrants; note that figures 22 and 23 are in dB. (The horizontal and vertical white streaks in figure 23 above 32 kHz should be ignored; they are due to a malfunction in producing the MATLAB .tif plots for inclusion in a Microsoft Word document. Several trials yielded these same results with streaks.)

It is not valid to compare the level of h_2 with the level of h_1 ; these two time-domain kernels don't even have the same dimensions; this can be seen from equation (1), where an extra x multiplication takes place at second order. The same comment applies to attempting to compare the level of frequency-domain kernel H_2 with the level of H_1 ; these two quantities don't have the same dimensions either. The only valid comparison of levels is in terms of the Volterra model output waveforms $y_1(n)$ and $y_2(n)$, which always *do* have the same dimensions and which are the same as those of $z(n)$.

A plot of a section of the fitted first-order waveform $y_1(n)$ is given in figure 24. It has no jagged behavior, as was present in microphone output $z(n)$ in figure 12 or in laser input $x(n)$ in figures 13 and 14. This is consistent with the estimated first-order frequency-domain kernel H_1 in figure 19, which has its major response below 15 kHz.

The corresponding spectrum $Y_1(f)$ of $y_1(n)$ is displayed in figure 25. Its major contribution is in the frequency band from 5 to 9 kHz; this compares favorably with the result for the first-order frequency-domain kernel H_1 in figure 19. It should be noted in figure 25 that the zero-frequency level is approximately the same as the major response level at 8 kHz. In contrast, in figure 19, the zero-frequency level is significantly below that at 8 kHz. This difference is due to the fact that the excitation $x(n)$ of laser L_2 has a much higher spectral level near zero frequency than at 8 kHz, as can be seen in figure 16.

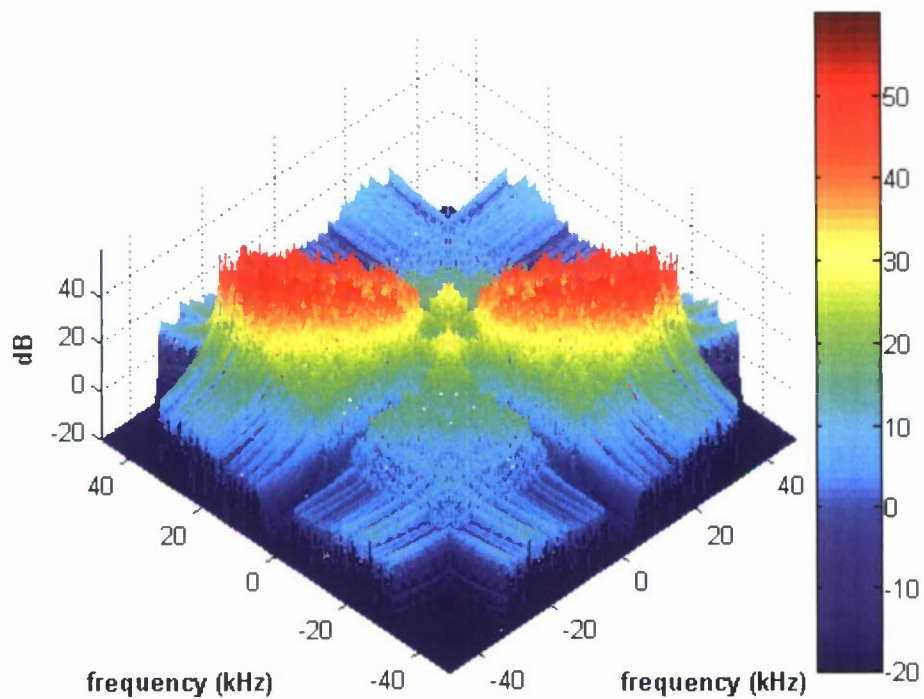


Figure 22. Second-Order Kernel H_2 (Observation Elevation Angle 50°)

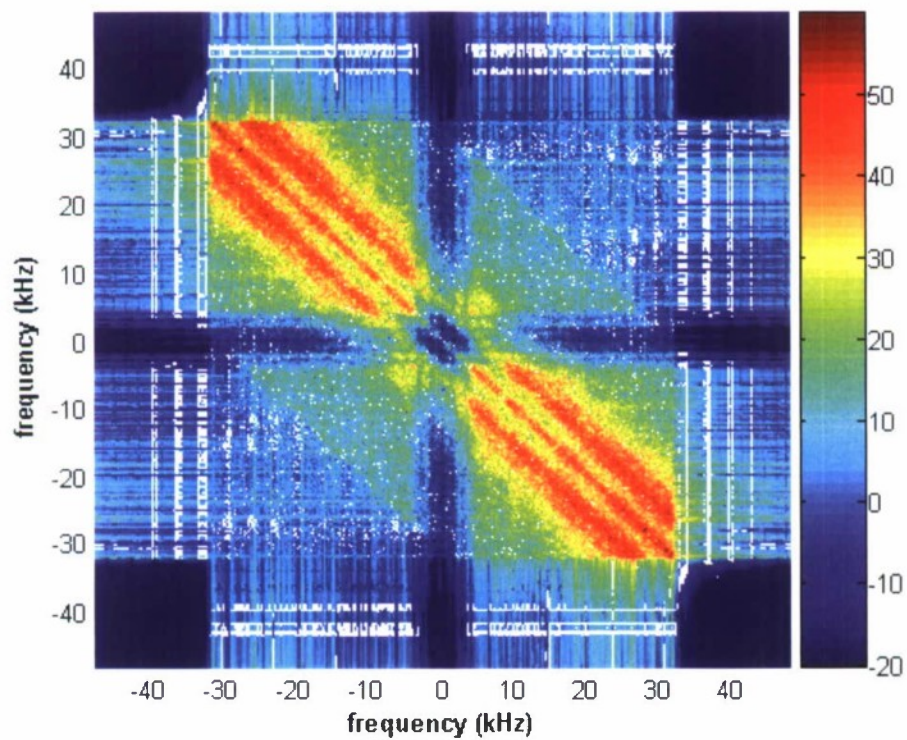


Figure 23. Second-Order Kernel H_2 (Observation Elevation Angle 90°)

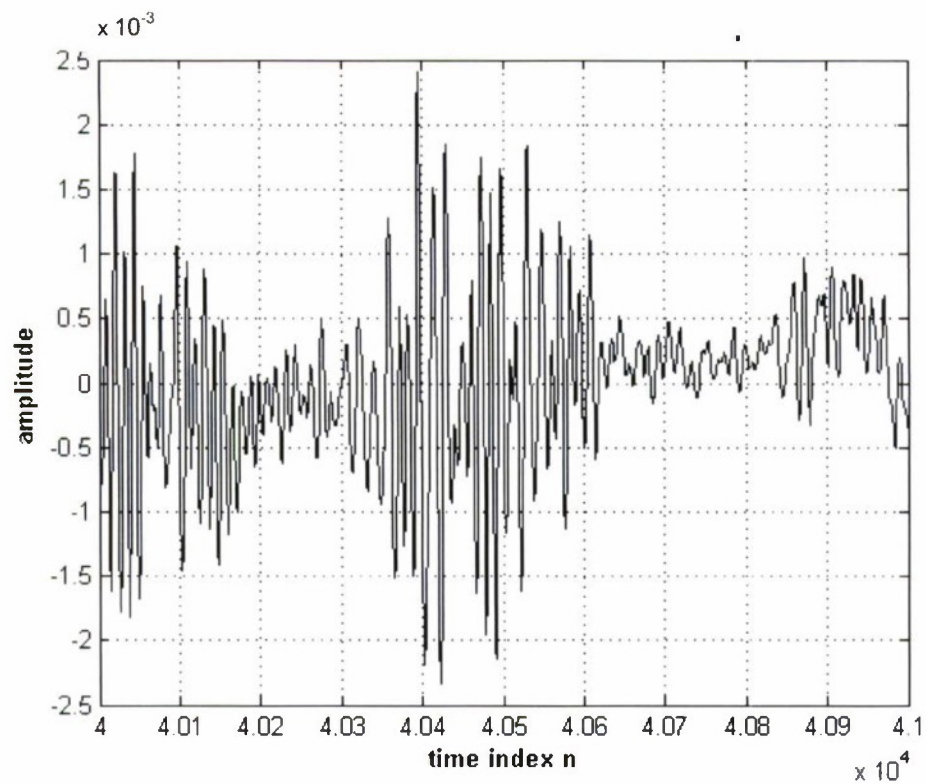


Figure 24. First-Order Waveform y_1

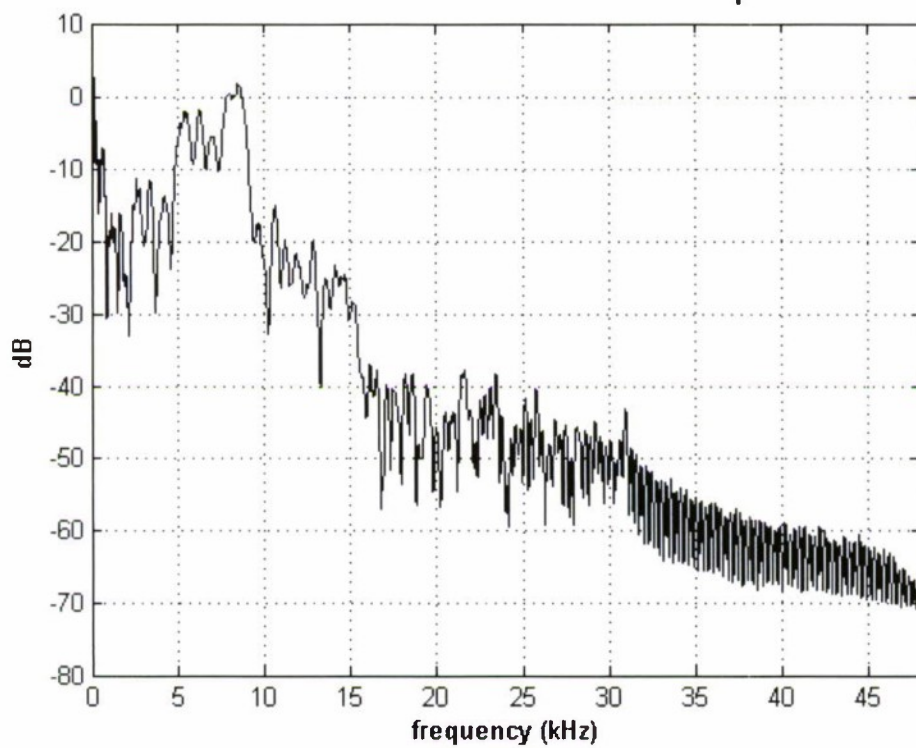


Figure 25. First-Order Spectrum Y_1

A section of second-order Volterra component waveform $y_2(n)$ is illustrated in figure 26. It has a somewhat more jagged behavior than first-order waveform $y_1(n)$ in figure 24. The corresponding spectrum $Y_2(f)$ of $y_2(n)$ is given in figure 27. A pronounced spectral contribution of Y_2 is centered at 6 kHz with a peak value of 7 dB. The first-order spectrum Y_1 in figure 25 has a peak value of only -3 dB. This 10-dB difference is due to the second-order frequency-domain kernel results for H_2 in figures 22 and 23. Namely, it is the ± 6 -kHz ridges of H_2 in figure 23 that are creating difference frequencies in the neighborhood of 6 kHz. Here, it is valid to make these quantitative comparisons of levels between Y_1 and Y_2 , because the same spectral estimation techniques and parameters were used on $y_1(n)$ and $y_2(n)$, which have the same dimensions.

The corresponding section of total Volterra fit $y(n)$ is plotted (in red) in figure 28, superposed on the measured microphone response $z(n)$ (in black). As can be seen, there are time segments where the total fit is especially good, and other time segments where the fit is poor. This is partially attributed to the fact that only laser L_2 was used for this least squares fit. There is the additional input from laser L_1 that was totally ignored during this single-input least squares fit.

All the fitting results thus far have been for laser L_2 of cicada song S14 considered as the input. For the fitting results in figure 29, laser L_1 was used instead as the input $x(n)$. The measured microphone output $z(n)$ is plotted in black, fit $y_1(n)$ is plotted in blue, and fit $y_2(n)$ is plotted in red. For this segment of time, the first-order fit is extremely small, while the second-order fit does a decent job of mimicking output $z(n)$. In fact, for this particular time segment, the ratio of powers in $y_1(n)$ to $z(n)$ is 0.036, while the ratio of powers in $y_2(n)$ to $z(n)$ is 0.68. This example, along with figure 28, brings out the point that the cicada with two laser measurements and one microphone measurement should be considered as a *two*-input, one-output nonlinear system with memory. That is, the block diagram in figure 1 is not general enough to encompass this generalization. More will be said on this observation below.

Results for a different cicada example (namely, song S9) are given in figures 30 and 31 for lasers L_1 and L_2 , respectively. Again, the measured microphone output $z(n)$ is plotted in black, fit $y_1(n)$ is plotted in blue, and fit $y_2(n)$ is plotted in red. The fractional power fit of total waveform $y(n)$ to output $z(n)$ in figure 30 for laser L_1 is 0.25 at the left edge of the figure, but 0.92 at the right edge of the figure. By contrast, the fractional power fit of total waveform $y(n)$ to output $z(n)$ in figure 31 for laser L_2 is 0.78 at the left edge of the figure, but 0.24 at the right edge of the figure. This example serves to further support the fact that different tymbals (lasers) are contributing to different portions of the time waveform emitted from the cicada. The only way to properly calculate the relative strengths of these components is by way of a two-input, one-output Volterra model.

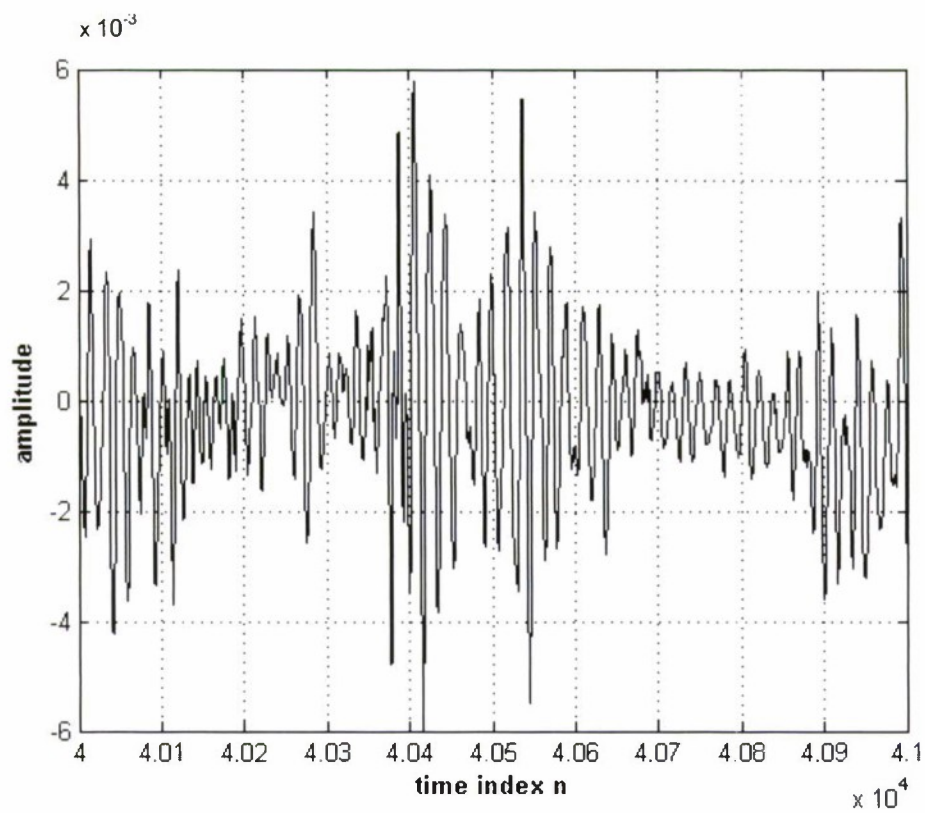


Figure 26. Second-Order Waveform y_2

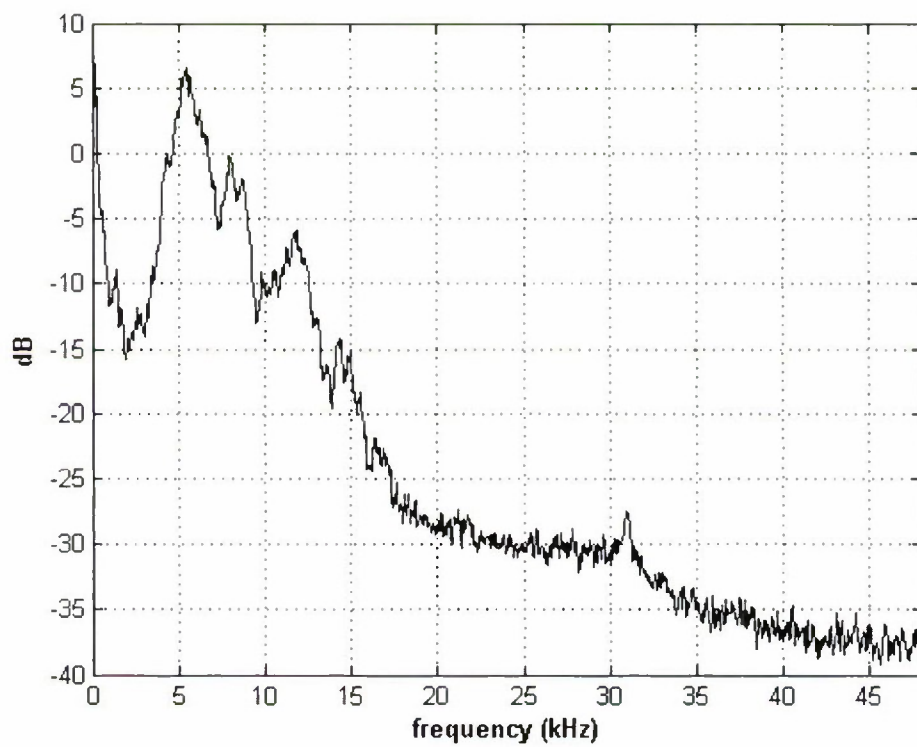


Figure 27. Second-Order Spectrum Y_2

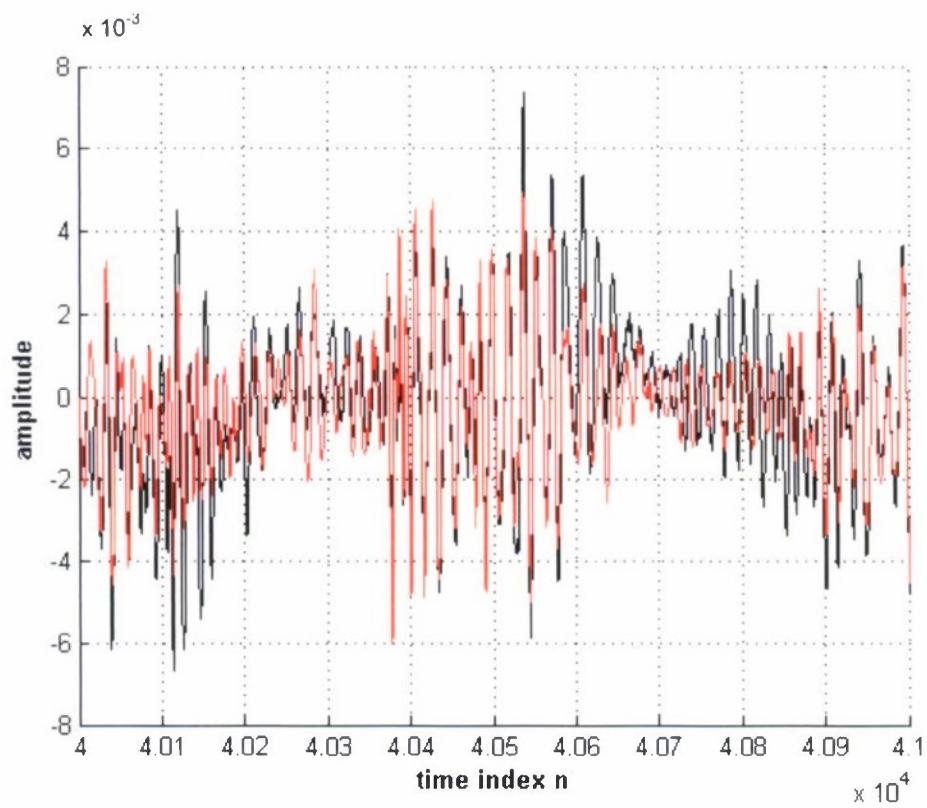


Figure 28. Total Waveforms $z(k)$ $y(r)$

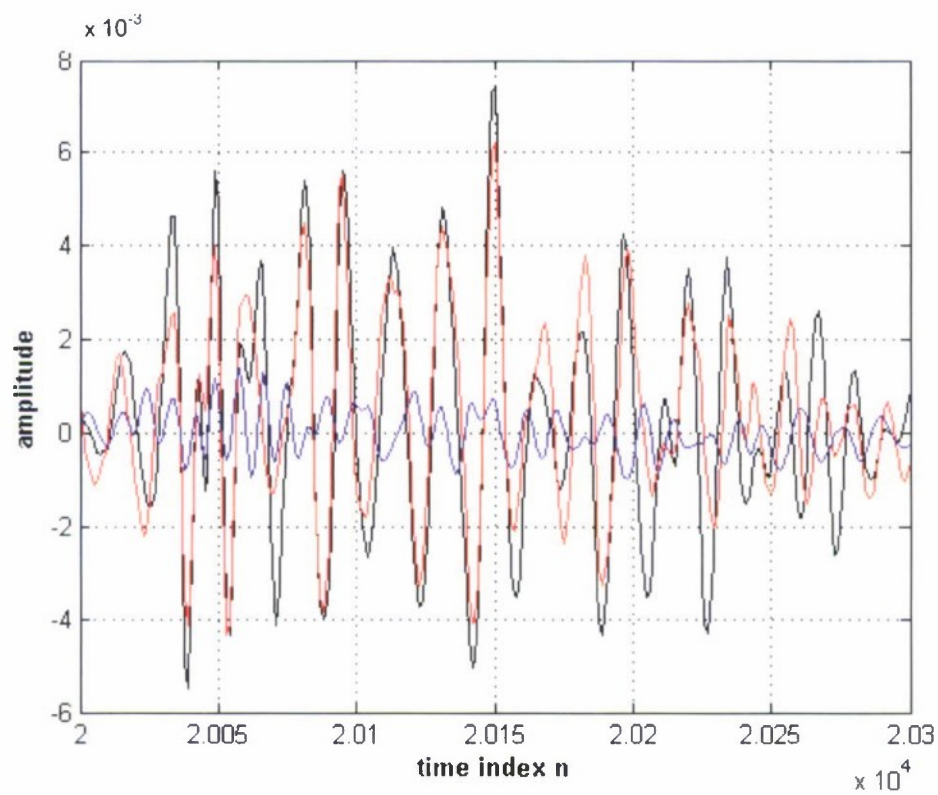


Figure 29. Waveforms $z(k)$ $y_1(b)$ $y_2(r)$ (Laser L_2 of Cicada Song S14)

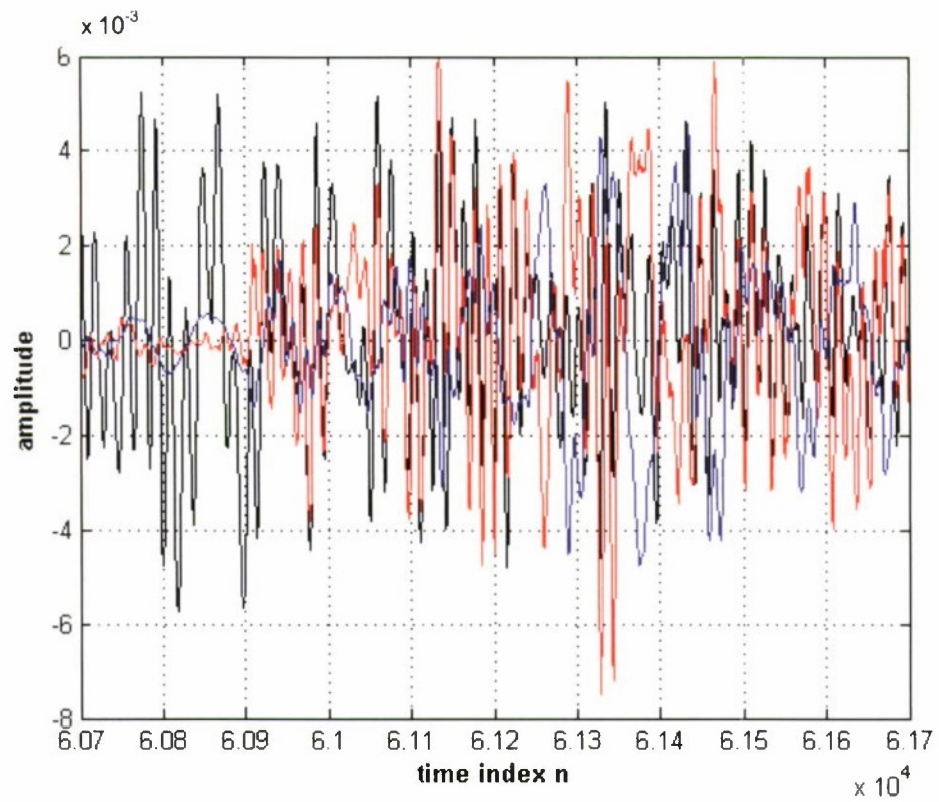


Figure 30. Waveforms $z(k)$ $y_1(b)$ $y_2(r)$ (Laser L_1 of Cicada Song S9)

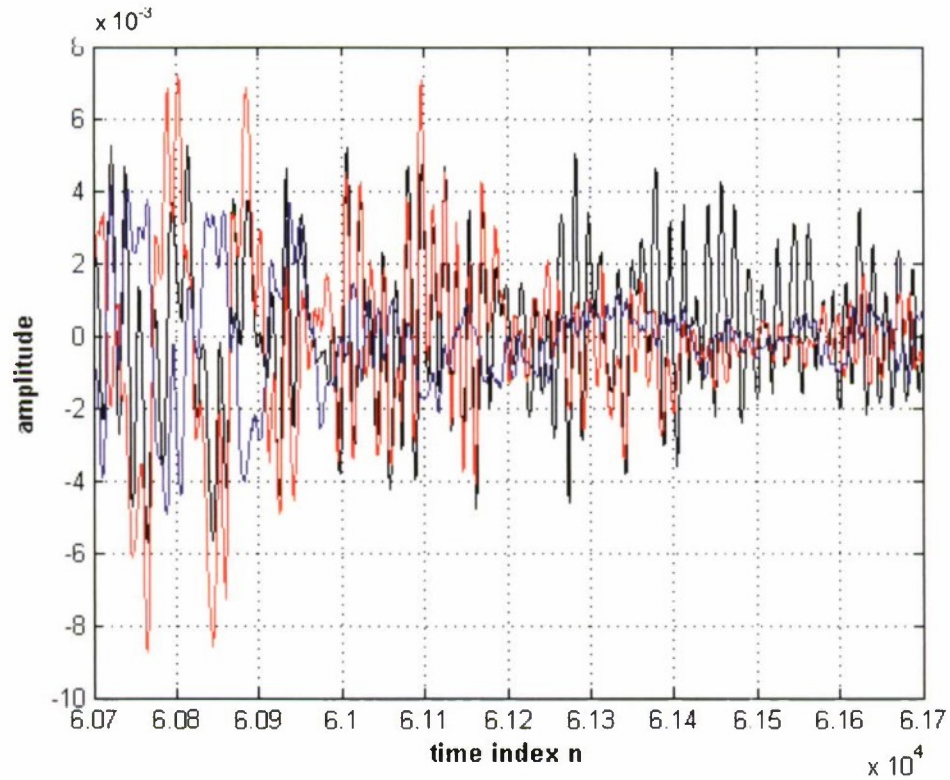


Figure 31. Waveforms $z(k)$ $y_1(b)$ $y_2(r)$ (Laser L_2 of Cicada Song S9)

8. SECOND-ORDER, TWO-INPUT, ONE-OUTPUT VOLTERRA MODEL

A Volterra model of a second-order, two-input, one-output system is given in figure 32. The two inputs are now labeled $x_1(n)$ and $x_2(n)$. There are now five required kernels that must all be estimated from simultaneous recordings of laser inputs $x_1(n)$ and $x_2(n)$ and microphone response $z(n)$. The number of subscripts on the kernels h denotes their order, either first order or second order. A new type of kernel must now be employed, namely, a "cross-kernel" $h_{12}(k_1, k_2)$, which has *two* inputs and one output and which yields a second-order waveform

$$y_{12}(n) = \sum_{k_1=0}^{K-1} \sum_{k_2=0}^{K-1} h_{12}(k_1, k_2) x_1(n-k_1) x_2(n-k_2) \quad (76)$$

$$= \Delta^2 \iint df_1 df_2 \exp[i2\pi(f_1 + f_2)n\Delta] H_{12}(f_1, f_2) X_1(f_1) X_2(f_2).$$

However, time-domain cross-kernel $h_{12}(k_1, k_2)$ need *not* be symmetric, because two different inputs x are multiplied in the first line of equation (76). Therefore, the only symmetry relation satisfied by its corresponding second-order frequency-domain kernel is

$$H_{12}(-f_1, -f_2) = H_{12}(f_1, f_2)^*, \quad (77)$$

which follows from the imposed realness of $h_{12}(k_1, k_2)$.

If Volterra output $y_{12}(n)$ is to have frequency content only in the band (f_a, f_b) , for purposes of fitting to a band-limited version of $z(n)$, and if $X_1(f)$ and $X_2(f)$ are broadband, the second line of equation (76) reveals that one must have frequency-domain kernel $H_{12}(f_1, f_2)$ nonzero only for

$$f_a < |f_1 + f_2| < f_b, \quad (78)$$

just as in equation (34). Therefore, it will be necessary to let variables f_1, f_2 in figure 2 range not only over the blue region, but, in addition, over its green extension into the second quadrant. The number of unknown (complex) coefficients for H_{12} will be equal to the sum of the number of coefficients required for each of H_{11} and H_{22} . This larger number of coefficients will significantly increase the effort required in the least squares procedure.

The basis functions for the $h_{12}(k_1\Delta, k_2\Delta)$ kernel are given by one-half of equation (36) and its conjugate (but not using the symmetry property in the top line of equation (37)), namely,

$$a(m_1, m_2) \cos(A) + b(m_1, m_2) \sin(A), \quad (79)$$

where A is given by equation (38). The $b(m_1, -m_1)$ terms must again be dropped from the basis set.

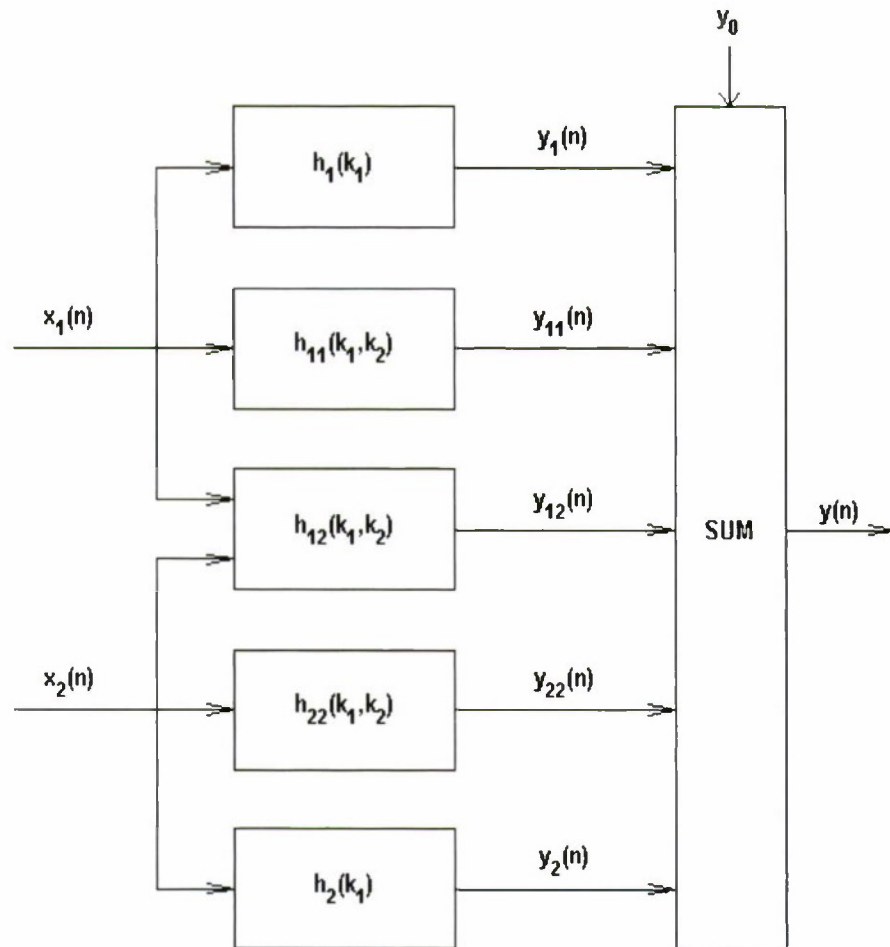


Figure 32. Second-Order, Two-Input, One-Output Volterra Model

9. SUMMARY

Several new concepts have been introduced to achieve some alleviation of the curse of dimensionality in determining the higher order kernels in a Volterra expansion. These include: partitioning of the frequency scale(s), the fundamental region of formation, the diagonal strips in the two-dimensional frequency space, the second- and third-order basis functions, and filtering of the measured nonlinear output to the current frequency band under investigation. The equations for the real basis functions at second and third order take some unexpected forms that could probably not be anticipated. These results were based on taking full advantage of symmetry and conjugate-symmetry frequency-domain relations that exist for real, symmetric, second-order and third-order, time-domain kernels. The inability to construct ideal bandpass filters requires the use of a frequency-overlap procedure, followed by the discarding of edge estimates with inherent errors, and the retention of only the interior estimates of higher accuracy.

Application to a first- and second-order (noise-free) control example with a white broadband excitation gave excellent estimates of all the first- and second-order properties, such as the individual kernels, the individual Volterra waveforms, and the total estimated output waveform. Application to a cicada mating call with a distinctly non-white and non-Gaussian excitation gave good results for the estimated first- and second-order kernels and waveforms, considering the non-optimality of this type of excitation. However, if the design of the input excitation to a nonlinear system with memory is under the user's control, the ideal excitation to use is a white, Gaussian process. Better correlation properties ensue and smaller condition numbers prevail in this latter situation.

The reduction factor in the number of coefficients to be determined at each frequency partition is given by W/F , where W is the bandwidth of the frequency interval under investigation, and F is the largest frequency of interest, typically near the Nyquist frequency. This factor applies at all orders and is *helpful* in alleviating the curse of dimensionality, but does not eliminate it.

INITIAL DISTRIBUTION LIST

Addressee	No. of Copies
Defense Advanced Research Projects Agency (Attn: D. Furey)	1
Center for Naval Analyses	1
Defense Technical Information Center	2
Albert Nuttall, Old Lyme, CT	10
Duncan Sheldon, Portsmouth, RI	1
Martin Schetzen, Brookline, MA	1